

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:22:05 ; Search time 1729.86 Seconds
(without alignments)
1153.942 Million cell updates/sec

Title: US-09-904-420A-1_COPY_820_940

Perfect score: 121

Sequence: 1 accacggcatgtacaaaga.....gctggttaagtgaaccccgcc 121

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	121	100.0	2855	9	HSTCF1G	X63901 Homo sapien
2	120	99.2	84544	2	AC009012	AC009012 Homo sapi
3	120	99.2	86914	2	AC011336	AC011336 Homo sapi
4	120	99.2	159420	2	AC009017	AC009017 Homo sapi
5	120	99.2	200831	2	AC008608	AC008608 Homo sapi
6	105	86.8	1254	9	HSTCF1A	X59869 Human TCF-1
7	105	86.8	2814	9	HSTCF1B	X59870 Human TCF-1
8	105	86.8	2910	9	HSTCF1C	X59871 Human TCF-1
9	97	80.2	1115	9	HSTCF1D	Z47361 H.sapiens T
10	97	80.2	1165	9	HSTCF1E	Z47362 H.sapiens T
11	91.8	75.9	162496	2	AC044846	AC044846 Mus muscu
12	88	72.7	1790	10	MMTCSYF	X61385 Mouse mRNA
13	34.8	28.8	166523	2	AC022646	AC022646 Homo sapi
14	34.8	28.8	198917	9	AL356157	AL356157 Human DNA
15	34.6	28.6	159135	2	AC091039	AC091039 Homo sapi
16	33.2	27.4	156555	9	AC026756	AC026756 Homo sapi
17	33.2	27.4	182507	2	AL356486	AL356486 Homo sapi
18	33	27.3	360	9	HUMCMP7	M5681 Human carti
19	32.6	26.9	38789	2	AC008145	AC008145 Homo sapi
20	32.6	26.9	191601	2	AC092143	AC092143 Homo sapi
21	32.4	26.8	5187	1	AB039821	AB039821 Gluconob
22	32.2	26.6	298	9	HSTGFBRI1G	AF054590 Homo sapi
23	32	26.4	319	9	HSB83F04N	X87064 H.sapiens B
24	32	26.4	441	9	S55290	S55290 anti-tetanu
25	32	26.4	2391	4	BCVDOPATK	M80234 Cow dopamin
26	32	26.4	3918	4	BTCPOLAPII	X89969 H.Laurus mR
27	31.8	26.3	3043	14	AF226637	AF226637 Cercopith
28	31.8	26.3	52173	9	HSG6PDGEN	X55448 H.sapiens G
29	31.8	26.3	160656	2	AL592072	AL592072 Homo sapi
30	31.8	26.3	215961	2	AC012363	AC012363 Homo sapi
31	31.8	26.3	219447	9	HOMFLNGEPD	I44140 Homo sapien
32	31.6	26.1	260	9	AF103166	AF103166 Homo sapi
33	31.6	26.1	179376	2	AC022809	AC022809 Homo sapi
34	31.4	26.0	140364	2	AL137027	AL137027 Homo sapi
35	31.4	26.0	152547	2	AL137857	AC079140 Homo sapi
36	31.4	26.0	158203	2	AC079140	AC079140 Homo sapi
37	31.4	26.0	166683	2	AC078821	AC078821 Homo sapi
38	31.2	25.8	35465	9	AC005525	AC005525 Homo sapi
39	31.2	25.8	38611	9	AC011545	AC011545 Homo sapi
40	31.2	25.8	108061	8	NC104H10	AL513410 Neurospor
41	31.2	25.8	140606	2	AL353761	AL353761 Homo sapi
42	31.2	25.8	156506	2	AC068038	AC068038 Homo sapi
43	31.2	25.8	162984	2	AL590642	AL590642 Homo sapi
44	31.2	25.8	198266	2	AL356156	AL356156 Homo sapi
45	31.2	25.8	214821	2	AC068785	AC068785 Homo sapi

ALIGNMENTS

RESULT 1

HSTCF1G

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

30-SEP-1999

HSTCF1G 2855 bp DNA

Homo sapiens TCF-1 gene.

X63901

X63901.1 GI:36791

T-cell transcription factor; transcription factor.

human.

Homo sapiens

Eukaryota; Metazoa;

Mammalia; Futheria;

1 (bases 1 to 2855)

van de Wetering,M.,

Suijkerbuijk,R.,

The human T cell

localization, and promoter

J. Biol. Chem. 267 (12),

92235082

2 (bases 1 to 2855)

Van de Wetering,M.,

Castrop,J.,

Korinek,V. and

Clevers,H.

Thu Feb 21 07:25:34 2002

Extensive alternative splicing and dual promoter usage generate
Tcf-1 protein isoforms with differential transcription control

Mol. Cell. Biol. 16 (3), 745-752 (1996)

96182076

3 (bases 1 to 2855)

Van de Wetering, M.L.

Direct Submission

Submitted (07-JAN-1992) M.L. Van de Wetering, Department of

Immunology, P.O. Box 8550, 3508 GA Utrecht, THE NETHERLANDS

Location/Qualifiers

1. 2855

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/cell_type="T-Lymphocyte"

/clone="phi-TCF-[5,8,11]"

/map="5q31.1"

829..924

/gene="TCF-1"

/number=1

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

/codon_start=1

/protein_id="CAB56795.1"

/db_xref="GI:6006565"

/translation="MYKETVYSAFNLIMHYPPPGACQHPQPPPIIKANQPPRGVPO

LSLIEFNSPHPTAPADISQKVHRPILOTPLDSGFSLTSGMGLPHIVSNFTHPS

LMLGSGVGPHPAAIPHPAIVPPSGKQELQDFDLNLTQAESKAKEAKPTIKRPLNA

FMLYMKEMRAKVIACETLKASAAIINQILGRWRHALSREQAKYVELARKERQLHMOLY

PGMSARDNYGKKRRSRREKHQESTTETINPRELDKNGQESLSMSSSSSPA"

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

925..970

/number=1

970..1074

/gene="TCF-1"

/number=2

1075..1119

/number=2

1120..1209

/gene="TCF-1"

/number=3

1210..1254

/number=3

1255..1374

/gene="TCF-1"

/number=4

1375..1419

/number=4

1420..1581

/gene="TCF-1"

/number=5

1582..1690

/number=5

1691..1798

/gene="TCF-1"

/number=6

1799..1842

/number=6

1843..1890

/gene="TCF-1"

/number=7

1891..2053

/number=1

2054..2134

/gene="TCF-1"

/number=8

592 a 929 c 771 g 562 t 1 others

BASE COUNT

ORIGIN

Query Match 100.0% Score 121: 108 9: Length 2855;
Best Local Similarity 100.0% Pred. No. 3.2e-23; Mismatches 0; Indels 0; Gaps 0;
Matches 121; Conservative 0;

Qy 1 accagcgcatgtacaaagagacggtctactccgccttcaatctgctcatgcatcacc 60
|||||
Db 820 ACCAAGCGCATGTACAAAGAGACCGTCTACTCCGCTTCAATCTGCTCATGCAATACCCA 879
|||||
Qy 61 ccccccctggagcagggcagcaccacccagcagcagcccccgcgtatgaagtgagaccgc 120
|||||
Db 880 CCCCCCTCGGAGCAGGCGCAGCCGCCAGCCGCCGCTGCTGTAAGTGGAGCCCGC 939
|||||
Qy 121 c 121
Db 940 C 940

RESULT 2
AC009012/c 84544 bp DNA HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone XXpl-360D11, WORKING DRAFT
DEFINITION AC009012, 7 ordered pieces.
ACCESSION AC009012.3 GI:12830104
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 84544)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 84544)
REFERENCE Direct Submission
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 14, 2001 this sequence version replaced gi:7454202.

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center project Name: 1189075, H71
Center clone name: XXpl-360D11

Summary Statistics
Consensus quality: 78189 bases at least Q40
Consensus quality: 82001 bases at least Q30
Consensus quality: 83201 bases at least Q20
Estimated insert size: 8000; pulse field gel estimation
Estimated insert size: 83944; sum-of-ctnigs estimation
Quality coverage: 6.34 in Q20 bases; pulse field gel estimation
Quality coverage: 6.04 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2979: contig of 2979 bp in length
* 2980 3079: gap of unknown length
* 3080 34576: contig of 31497 bp in length
* 34577 34676: gap of unknown length
* 34677 59262: contig of 24586 bp in length
* 59263 59362: gap of unknown length
* 59363 76010: contig of 16648 bp in length

* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86914: contig of 86914 bp in length.
Location/Qualifiers
1. 86914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-250113"
/clone_lib="Caltech human BAC library C"
BASE COUNT 20745 a 22135 c 22612 g 21422 t
ORIGIN

Query Match 99.2%; Score 120; DB 2; Length 86914;
Best Local Similarity 100.0%; Pred. No. 3.5e-23;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 accagggcatgtacaaagagacgcttactcgccttcaatctgctcatgtaccaca 60
|||||
Db 45323 ACAGGGCATGTCAAAGAGACGCTTACTCGGCTTCAATCTGCTCATGTATACCA 45264
|||||

QY 61 cccctcggagagggcagcagcccccagcagcccccgcagcccccgtgtaagtgcacccgc 120
|||||
Db 45263 CCGGCTCGGAGCAGGGCAGCACCCCGCAGCCGCGCGGCTGCTAAGTGGACCCCGC 45204
|||||

RESULT 4
AC009017/C
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone XXpl-929G6, WORKING DRAFT SEQUENCE,
26 unordered pieces.
AC009017
VERSION
GI:13699469
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 159420)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 159420)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:11178048.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1189133, H51
Center clone name: XXpl-929G6

Summary Statistics
Consensus quality: 142845 bases at least Q40
Consensus quality: 150910 bases at least Q30
Consensus quality: 152756 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Quality coverage: 8.41 in Q20 bases; sum-of-contigs estimation
Quality coverage: 4.29 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* 76011 76110: gap of unknown length
* 76011 79613: contig of 3503 bp in length
* 79613 79713: gap of unknown length
* 79713 80649: contig of 936 bp in length
* 80649 80749: gap of unknown length
* 80749 84544: contig of 3795 bp in length.
Location/Qualifiers
1. 84544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXpl-360D11"
BASE COUNT 19558 a 21740 c 22203 g 20376 t 667 others
ORIGIN

Query Match 99.2%; Score 120; DB 2; Length 84544;
Best Local Similarity 100.0%; Pred. No. 3.5e-23;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 accagggcatgtacaaagagacgcttactcgccttcaatctgctcatgtaccaca 60
|||||
Db 38748 ACCAGGGCATGTCAAAGAGACGCTTACTCGGCTTCAATCTGCTCATGTATACCA 38689
|||||

QY 61 cccctcggagagggcagcagcccccagcagcccccgcagcccccgtgtaagtgcacccgc 120
|||||
Db 38688 CCGGCTCGGAGCAGGGCAGCACCCCGCAGCCGCGCGGCTGCTAAGTGGACCCCGC 38629
|||||

RESULT 3
AC011336/C
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone CTC-250113, WORKING DRAFT SEQUENCE,
1 ordered pieces.
AC011336
VERSION
GI:9256274
KEYWORDS
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEPIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86914)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 86914)
DOE Joint Genome Institute.
Direct Submission
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710593.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 281384, H247
Center clone name: C1F-HSPC_250113

Summary Statistics
Consensus quality: 82614 bases at least Q40
Consensus quality: 85602 bases at least Q30
Consensus quality: 86382 bases at least Q20
Estimated insert size: 89000; pulse field gel estimation
Quality coverage: 86914; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 6.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1423: contig of 1423 bp in length
1424 1523: gap of unknown length
1524 3066: contig of 1543 bp in length
3067 3166: gap of unknown length
3167 5629: contig of 2463 bp in length
5630 5729: gap of unknown length
5730 6997: contig of 1268 bp in length
6998 7097: gap of unknown length
7099 9214: contig of 2117 bp in length
9215 9315: gap of unknown length
9316 11813: contig of 2499 bp in length
11814 14718: contig of 2805 bp in length
14719 14818: gap of unknown length
14819 17174: contig of 2356 bp in length
17175 17274: gap of unknown length
17275 19710: contig of 2436 bp in length
19711 21678: gap of unknown length
21679 21778: gap of unknown length
21779 26021: contig of 4243 bp in length
26022 26121: gap of unknown length
26122 29609: contig of 3488 bp in length
29610 29709: gap of unknown length
29710 33196: contig of 3487 bp in length
33197 33296: gap of unknown length
33297 36995: contig of 3699 bp in length
36996 37095: gap of unknown length
37096 40646: contig of 3551 bp in length
40647 40746: gap of unknown length
40747 45044: contig of 4298 bp in length
45045 45144: gap of unknown length
45145 52148: contig of 7004 bp in length
52149 52248: gap of unknown length
52249 57686: contig of 5438 bp in length
57687 57786: gap of unknown length
57787 66113: contig of 8327 bp in length
66114 66213: gap of unknown length
66214 75538: contig of 9325 bp in length
75539 75638: gap of unknown length
75639 82249: contig of 6611 bp in length
82250 82349: gap of unknown length
82350 95465: contig of 13116 bp in length
95466 110942: gap of unknown length
110943 111042: contig of 15377 bp in length
111043 127290: gap of unknown length
127291 127390: gap of unknown length
127391 143681: contig of 16291 bp in length
143682 143782: gap of unknown length
143782 159420: contig of 15639 bp in length.

FEATURES
source
1..159420
/organism="homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXpl-929G6"

BASE COUNT 39367 a 38414 c 39218 g 39915 t 2506 others
ORIGIN

Query Match 99.2%; Score 120; DB 2: length 159420;
ResL Local Similarity 100.0%; Pred. No. 3.2e-23;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 accagcgcatgtacaagagacgctctactccgcctcaactctgtcatgcatatccca 60
|||||
DB 56877 ACCAGCGCATGTACAAGAGACCGCTCTACTCCGCTTCAATCTCTCATGATTTACCCA 56818
|||||
QY 61 cttccctcgggagcaggcagcagcccccagccagccccccgctgtgaattttatcttcc 120

|||||
Db 56817 CCCCCTCGGAGCAGGCGACACCCCGAGCCCGCTGTAGTGGACCCCGC 56758
|||||
RESULT 5
AC008608 200831 bp DNA HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTB-113120, WORKING DRAFT SEQUENCE,
DEFINITION 29 unordered pieces.
AC008608
AC008608.5 GI:13699394
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 200831)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 200831)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE On Apr 20, 2001 this sequence version replaced qi:7708985.
JOURNAL
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center project Name: 107808, H233
Center clone name: C107808KB_113120

Summary Statistics
Consensus quality: 174733 bases at least Q40
Consensus quality: 186693 bases at least Q30
Consensus quality: 190547 bases at least Q20
Estimated insert size: 165000; pulse field gel estimation
Estimated insert size: 198031; sum-of-contigs estimation
Quality coverage: 7.77 in 020 bases; pulse field gel estimation
Quality coverage: 6.47 in 020 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1115: contig of 1115 bp in length
* 1116 1215: gap of unknown length
* 1216 2567: contig of 1352 bp in length
* 2568 2667: gap of unknown length
* 2668 3745: contig of 1078 bp in length
* 3746 3845: gap of unknown length
* 3846 4920: contig of 1075 bp in length
* 4921 5020: gap of unknown length
* 5021 6044: contig of 1024 bp in length
* 6045 6144: gap of unknown length
* 6145 7285: contig of 1141 bp in length
* 7286 7386: gap of unknown length
* 7387 8659: contig of 1274 bp in length
* 8660 8759: gap of unknown length
* 8760 9804: contig of 1045 bp in length
* 9805 9904: gap of unknown length
* 9905 11696: contig of 1792 bp in length
* 11697 11797: gap of unknown length
* 11798 12841: contig of 1045 bp in length
* 12842 12942: gap of unknown length
* 12943 15043: contig of 2102 bp in length
* 15044 15143: gap of unknown length

FEATURES	source
BASE COUNT	53266 a 46907 c 47140 g 50695 t 2823 others
ORIGIN	
Query Match	99.28; Score 120; DB 2; Length 200831;
Best Local Similarity	100.0%; Pred.No. 3; Ie-23;
Matches 120; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 1	accagcgcatatacaagagacgctactccgccttaactgctcalgrattaccra 60
Db 183977	ACCAGCGGCATCTACAAAGACGCGTCTACTCCGCTTCAATCTGCTCATGGCATTACCCA 184036
QY 61	ccccctcgagagcaggcagcagcccccagccagcccgctglaaataagacccac 120
Db 184037	CCCCCTCGGAGCAGGACAGCCGCCAGCCGCGTGTGTAAGTATGACCCGCG 184096
RESULT 6	
HSTCFIA	1254 bp mRNA
LOCUS	Human TCF-1 mRNA for T cell factor 1 (splice form A)
DEFINITION	X59869.1 X55327
ACCESSION	X59869.1 GI:36785
KEYWORDS	DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1254) van de Wetering, M. Direct Submission.

Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

REFERENCE 2 (bases 1 to 2814)
AUTHORS van de Wetering,M., Oosterwegel,M., Doolies,D. and Clevers,H.
TITLE Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box

JOURNAL EMBO J. 10 (1), 123-132 (1991)

MEDLINE 91114695

COMMENT See also X59869-X59871.

FEATURES Location/Qualifiers

source 1..2814

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="T lymphocyte"

/cell_line="Jurkat and HPB-ALL"

/clone_lib="cDNA"

mRNA 1..2814

/gene="TCF-1"

/evidence=experimental

gene 1..2814

/gene="TCF-1"

80..889

/gene="TCF-1"

/codon_start=1

/product="T cell factor 1 splice form B"

/protein_id="CAA42527.1"

/db_xref="GI:36788"

/translation="MYKETVVSAPNLLMHYPSPGAGQHPOPLHKANQPHGVPO

LSLVHFNSPHTPADISQKVHRLQTPDLISGFYSLSGSMGLPHTVSWFTHPS

LMLGSGVCHPAAIPHPAIVPPSGKQLQPDRLNLTQAEKAEKAKPTIKKPLNA

PMLYKMKRAKVIACFTLKESAAINQILGRWHALSRQDAKYIELAKRKRLHMLY

PCWSARDNYGKKRRSRKHKQESTTGKRNAGCTYPEKAAAPAPFLPMTVL"

misc_feature 539..769

/gene="TCF-1"

/note="HMG box"

misc_feature 810

/gene="TCF-1"

/note="alternative splice site"

BASE COUNT 763 a 831 c 648 g 572 t

ORIGIN

Query Match 86.8%; Score 105; DB 9; Length 2814;
Best Local Similarity 100.0%; Pred. No. 7.6e-19;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 accagcgcatgtacaaagagaccgtctactccgcttcaatctgctcatgcatlaccca 60

|||||

Db 71 ACCAGCGCATGTACAAAGAGACCGCTCTACTCCGCTTCAATCTGCTCATGCAATACCCA 140

Qy 61 cccccctcggaagcaggcagaccgccagccagcccccgcctg 105

|||||

Db 131 CCCCCCTCGGAGCAGGAGCACCACCCAGCCGCGCCGCTG 175

RESULT 8

LOCUS HSTCF1C

DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form C).

ACCESSION X59871 X55328

VERSION X59871.1 GI:36789

KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2910)

AUTHORS van de Wetering,M.

TITLE Direct Submission

COMMENT Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical

Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

The Netherlands

REFERENCE 2 (bases 1 to 2910)

AUTHORS van de Wetering,M., Oosterwegel,M., Doolies,D. and Clevers,H.

TITLE Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box

JOURNAL EMBO J. 10 (1), 123-132 (1991)

MEDLINE 91114695

COMMENT See also X59869-X59871.

FEATURES Location/Qualifiers

source 1..2910

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="T lymphocyte"

/cell_line="Jurkat and HPB-ALL"

/clone_lib="cDNA"

mRNA 1..2910

/gene="TCF-1"

/evidence=experimental

gene 1..2910

/gene="TCF-1"

80..886

/gene="TCF-1"

/codon_start=1

/product="T cell factor 1, splice form C"

/protein_id="CAA42528.1"

/db_xref="GI:36790"

/translation="MYKETVVSAPNLLMHYPSPGAGQHPOPLHKANQPHGVPO

LSLVHFNSPHTPADISQKVHRLQTPDLISGFYSLSGSMGLPHTVSWFTHPS

LMLGSGVCHPAAIPHPAIVPPSGKQLQPDRLNLTQAEKAEKAKPTIKKPLNA

PMLYKMKRAKVIACFTLKESAAINQILGRWHALSRQDAKYIELAKRKRLHMLY

PCWSARDNYGKKRRSRKHKQESTTDPGSPKCKARFGLNQUTDWGPGCR"

misc_feature 539..769

/gene="TCF-1"

/note="HMG box"

misc_feature 810

/gene="TCF-1"

/note="alternative splice site"

BASE COUNT 782 a 860 c 671 g 597 t

ORIGIN

Query Match 86.8%; Score 105; DB 9; Length 2910;
Best Local Similarity 100.0%; Pred. No. 7.9e-19;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 accagcgcatgtacaaagagaccgtctactccgcttcaatctgctcatgcatlaccca 60

|||||

Db 71 ACCAGCGCATGTACAAAGAGACCGCTCTACTCCGCTTCAATCTGCTCATGCAATACCCA 130

Qy 61 cccccctcggaagcaggcagaccgccagccagcccccgcctg 105

|||||

Db 131 CCCCCCTCGGAGCAGGAGCACCACCCAGCCGCGCGCTG 175

RESULT 9

LOCUS HSTCF1D

DEFINITION H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.

ACCESSION 247361

VERSION 247361.1 GI:619881

KEYWORDS splice form D; T cell factor 1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1115)

AUTHORS Mayer,K., Wolff,F., Clevers,H. and Ballhausen,W.G.

TITLE The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon

JOURNAL IXA

COMMENT Unpublished

REFERENCE 2 (bases 1 to 1115)

AUTHORS van de Wetering,M., Oosterwegel,M., Holsteyne,F., Booyes,D.,

Thu Feb 21 07:25:34 2002

novel isoforms due to alternative splicing and usage of a new exon

Suijkerbuijk.R., Geurts van Kessel.A. and Clevers.H.
The human T cell transcription factor-1 gene. Structure,
localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
92235082
3 (bases 1 to 1115)
Ballhausen.W.G.
Direct Submission
Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
Schwabachanlage 10, Erlangen, Germany, D-91054
Location/Qualifiers
source
1. .1115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1115
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1115
/gene="TCF-1"
2. .754
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form D"
/db_xref="GI:619882"
/translation="MYKTVYSAFNLLMHYPPSGAGOHPOPPHLKANOPPHGVPO
LSIYHFNPHPTAPADISQKQVHRPLOTDLGFSYLSGSMGOLPHTVSWFTHPS
LMIGSGVPGHPAAIPHPAIVPPSGKQELQPDRLNLTQAFSKAEKAKPTIKKPLNA
FMLYMKEMRAKVIACETLKESAAINQILGRWHALSREBUAKYELARKERQLHMOLY
PGWSARDNYGKKRRSRKHQESTTDPGSKKRCARFGLNOUTDMWGCRRKKK:IKY
PGLGRCPSVPDSI/SALGCGSPAPQDPSYIII.PRPTELLTSPARILHPOVSPL
LSASOPQGHPRPPAAPCAHRYSNRNLDRWPSRIHTPGRIQEPTP"
731. 751
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1D"
BASE COUNT 272 a 407 c 265 g 171 t
ORIGIN
Query Match 80.2%; Score 97; DB 9; Length 1115;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 catgtacaaagagaccgtctactccgcccttcgaatctgtcattgcattaccacccccc 68
DB 1 CATGTACAAAGAGACCGTCTACTCCGCCCTTCAATCTGCTCATGCATTACCCACCCCGCTC 60
QY 69 gggagcaggcagcagcccccagccgagcccccctg 105
DB 61 GGGAGCAGGCAGCACCCCGCAGCGAGCCCGCGCTG 97
RESULT 10
HSTCF1E 1165 bp mRNA PRI 09-JAN-1995
LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.
DEFINITION 247362
ACCESSION 247362
VERSION 1 GI:619883
KEYWORDS splice form E; T cell factor 1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1165)
AUTHORS Mayer.K., Wolff.E., Clevers.H. and Ballhausen.W.G.
TITLE The human high mobility group (HMG)-box transcription factor TCF-1:

novel isoforms due to alternative splicing and usage of a new exon
IXA
Unpublished
2 (bases 1 to 1165)
van de Wetering.M., Oosterwegel.M., Holstege.F., Dooyes.D.,
Suijkerbuijk.R., Geurts van Kessel.A. and Clevers.H.
The human T cell transcription factor-1 gene. Structure,
localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
92235082
3 (bases 1 to 1165)
Ballhausen.W.G.
Direct Submission
Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
Schwabachanlage 10, Erlangen, Germany, D-91054
Location/Qualifiers
source
1. .1165
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1165
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1165
/gene="TCF-1"
2. .1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKTVYSAFNLLMHYPPSGAGOHPOPPHLKANOPPHGVPO
LSIYHFNPHPTAPADISQKQVHRPLOTDLGFSYLSGSMGOLPHTVSWFTHPS
LMIGSGVPGHPAAIPHPAIVPPSGKQELQPDRLNLTQAFSKAEKAKPTIKKPLNA
FMLYMKEMRAKVIACETLKESAAINQILGRWHALSREBUAKYELARKERQLHMOLY
PGWSARDNYGKKRRSRKHQESTTDPGSKKRCARFGLNOUTDMWGCRRKKK:IKY
PGLGRCPSVPDSI/SALGCGSPAPQDPSYIII.PRPTELLTSPARILHPOVSPL
LSASOPQGHPRPPAAPCAHRYSNRNLDRWPSRIHTPGRIQEPTP"
803. 1141
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN
Query Match 80.2%; Score 97; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 catgtacaaagagaccgtctactccgcccttcgaatctgtcattgcattaccacccccc 68
DB 1 CATGTACAAAGAGACCGTCTACTCCGCCCTTCAATCTGCTCATGCATTACCCACCCCGCTC 60
QY 69 gggagcaggcagcagcccccagccgagcccccctg 105
DB 61 GGGAGCAGGCAGCACCCCGCAGCGAGCCCGCGCTG 97
RESULT 11
AC04846/C 162496 bp DNA HTG 22-JUL-2000
LOCUS AC04846
DEFINITION Mus musculus chromosome 11 clone RP23-46J7 map 11, WORKING DRAFT
SEQUENCE, 23 unordered pieces.
ACCESSION AC04846
VERSION AC04846.2 GI:9364533

KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE 1 (bases 1 to 162496)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162496)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczkzy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melidrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Olivier, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 22, 2000 this sequence version replaced gi:7543812.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9174
Center clone name: 46_J-7
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149639 bases at least Q40
Consensus quality: 156386 bases at least Q30
Consensus quality: 158956 bases at least Q20
Insert size: 164000; agarose-tp
Quality size: 160296; sum-of-ctgigs
Quality coverage: 3.5 in Q20 bases; aqarose-tp
Quality coverage: 3.6 in Q20 bases; sum-of-ctgigs

* NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 1373: contig of 1373 bp in length
* 1374 1473: gap of 100 bp
* 1474 2786: Contig of 1313 bp in length
* 2787 2886: gap of 100 bp
* 2887 4262: contig of 1376 bp in length
* 4263 4362: gap of 100 bp
* 4363 6417: contig of 2055 bp in length

FEATURES
source
1..162496
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-46J7"
/clone_lib="RPC1-23 Female Mouse BAC"
1..1373
/note="assembly_fragment"
1474..2786
/note="assembly_fragment"
2887..4262
/note="assembly_fragment"
4363..6417
/note="assembly_fragment"
6518..9412
/note="assembly_fragment"
9513..10790
/note="assembly_fragment"
clone_end:SP6
vector_side:right
10891..15080
/note="assembly_fragment"
15181..18824
/note="assembly_fragment"
18925..23129
/note="assembly_fragment"
23230..26859
/note="assembly_fragment"
26960..32575
/note="assembly_fragment"
32676..38781
/note="assembly_fragment"
38882..44764
/note="assembly_fragment"

6418 6517: gap of 100 bp
6518 9412: contig of 2895 bp in length
9413 9512: gap of 100 bp
9513 10790: contig of 1278 bp in length
10791 10890: gap of 100 bp
10891 15080: contig of 4190 bp in length
15081 15180: gap of 100 bp
15181 18824: contig of 3644 bp in length
18825 18924: gap of 100 bp
18925 23129: contig of 4205 bp in length
23130 23229: gap of 100 bp
23230 26859: contig of 3630 bp in length
26860 26959: gap of 100 bp
26960 32575: contig of 5616 bp in length
32576 32675: gap of 100 bp
32676 38781: contig of 6106 bp in length
38782 38881: gap of 100 bp
38882 44764: contig of 5883 bp in length
44765 44864: gap of 100 bp
44865 52115: contig of 7251 bp in length
52116 52215: gap of 100 bp
52216 59927: contig of 7712 bp in length
59928 60027: gap of 100 bp
60028 67115: contig of 7088 bp in length
67116 67215: gap of 100 bp
67216 75188: contig of 7973 bp in length
75189 75288: gap of 100 bp
75289 85641: contig of 10353 bp in length
85642 85741: gap of 100 bp
85742 97849: contig of 12108 bp in length
97850 97949: gap of 100 bp
97950 109572: contig of 11623 bp in length
109573 109672: gap of 100 bp
109673 122441: contig of 12769 bp in length
122442 122541: gap of 100 bp
122542 140243: contig of 17702 bp in length
140244 140343: gap of 100 bp
140344 162496: contig of 22153 bp in length.

```
/note="assembly_fragment"
44865..52115
/note="assembly_fragment"
52216..59927
/note="assembly_fragment"
60028..67115
/note="assembly_fragment"
67216..75188
/note="assembly_fragment"
75289..85641
/note="assembly_fragment"
85742..97849
/note="assembly_fragment"
97950..109572
/note="assembly_fragment"
109673..122441
/note="assembly_fragment"
122542..140243
/note="assembly_fragment
clone_end:T7
vector_side:left"
140344..162496
/note="assembly_fragment"
42470 a 37212 c 36404 g 44204 t 2206 others
ORIGIN

Query Match 75.98; Score 91.8; DB 2; Length 162496;
Best Local Similarity 89.78; Pred. No. 1.6e-15;
Matches 102; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 ccagcgcatgtacaaagagaccgtctactccgcttcaatctgctcatgcatcaccac 61
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62043 CCAGTGGCATGTACAAAGAGACTGTCTACTCTGCTTCATCTGCTCATGCCCTACCCAC 61984

QY 62 cccctcgggagcggcgagcaccgcccgagcccgcccgctggtgaagtgaacccgc 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61983 CGGCTCGGAGCAGCGGAGCATCCGAGCCTCAACCCCGCTGTAAGTGAACCGAAGC 61925

RESULT 12
MMTCSTF MMTCSTF 1790 bp mRNA ROD 02-SEP-1991
LOCUS Mouse mRNA for T-cell specific transcription factor.
ACCESSION X61385
VERSION X61385.1 GI:547760
KEYWORDS HMG box; T-cell specific transcription factor; transcription
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1790)
Oosterweel,M., van de Wetering,M., Dooljes,D., Klomp,I.,
Winoto,A., Georgopoulos,K., Meijlink,P. and Clevers,H.
Cloning of murine TCF-1, a T cell-specific transcription factor
interacting with functional motifs in the CD3-epsilon and T cell
receptor alpha enhancers
J. Exp. Med. 173 (5), 1133-1142 (1991)
MEDLINE 91217625
FEATURES
Location/Qualifiers
1..1790
/organism="Mus musculus"
/isolate="M29/M5"
/db_xref="taxon:10090"
/tissue_type="thymus"
/cell_type="T cell"
```

```
/clone_lib="lambda Zap cDNA"
1..1790
190..1101
/codon_start=1
/product="T-cell specific transcription factor"
/protein_id="CAA43658.1"
/db_xref="GI:54761"
/db_xref="MGD:MG1:98507"
/db_xref="SWISS-PROT:O00417"
/translation="MYKTYVSAPNLLMPYPASGAGOHPOPPHINKPGOPPHGV
QISPLYHFSPPHPTAPADISQOGVHRRLQTPDLSGFYSLTSGSMGOLPHVTWSPS
PPLYPLSPCGYRQHFPAPTAAPGAPYRPTHSLMIGSGVPGHPAPAIHPALIVRSSG
KOELQYDRNI.KTQAEKAEKAKKPVIKKPLNAFLMYLKMKEAKYVIAETLKESAAI
NOILGRWHALSREQAKYVELARKQLHMOLYPGWSARDNYTGKKKRSREKHQEST
TGCKRNAFTYPEKAAAPAPLPMTVL"
misc_feature 489..581
/note="ACT splice/exon M5"
761..984
misc_feature /note="HMG-box"
BASE COUNT 423 a 587 c 436 g 343 t 1 others
ORIGIN

Query Match 72.7%; Score 88; DB 10; Length 1790;
Best Local Similarity 83.3%; Pred. No. 3.6e-14;
Matches 100; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 ccaqcgcatgtacaaagagaccgtctactccgcttcaatctgctcatgcatcaccac 61
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 182 CCAGTGGCATGTACAAAGAGACTGTCTACTTGCCTTCAATCTGCTCATGCCCTACCCAC 241

QY 62 cccctcgggagcggcgagcaccgcccgagcccgcccgctggtgaagtgaacccgc 121
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 242 CGGCTCGGAGCAGCGGAGCATCCGAGCCTCAACCCCGCTGCATACCAAGCCCGGCC 301

RESULT 13
AC022646 166523 bp DNA HTG 10-MAR-2001
LOCUS Homo sapiens clone RP11-28C18, WORKING DRAFT SEQUENCE, 18 unordered
DEFINITION pieces.
ACCESSION AC022646
VERSION AC022646.3 GI:7249102
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 166523)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Bohuslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneslor,J.,
Ferrelira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hago,B., Heaford,A., Horton,B.,
Howland,J.C., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,K., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thumanni,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
TITLE
JOURNAL
```


repeat_region /note="LTR40a repeat: matches 2. .519 of consensus"
 47015. .47093
 repeat_region /note="MER91A repeat: matches 105. .186 of consensus"
 47500. .47917
 repeat_region /note="L2 repeat: matches 2309. .2746 of consensus"
 49126. .49349
 repeat_region /note="MER20 repeat: matches 1. .216 of consensus"
 49611. .49655
 repeat_region /note="MIR repeat: matches 219. .262 of consensus"
 49662. .50194
 repeat_region /note="L2 repeat: matches 2096. .2661 of consensus"
 51919. .52224
 repeat_region /note="AluSq repeat: matches 1. .306 of consensus"
 53562. .53818
 repeat_region /note="LMB8 repeat: matches 5919. .6167 of consensus"
 53817. .54341
 repeat_region /note="LMB6 repeat: matches 5614. .6140 of consensus"
 54343. .54633
 repeat_region /note="AluSq repeat: matches 1. .291 of consensus"
 56076. .56128
 repeat_region /note="L2 repeat: matches 2662. .2709 of consensus"
 complement(57923. .58306)
 /note="match: GSS: Em:AQ146681"
 /note="match: GSS: Em:AQ146681"
 complement(58138. .58635)
 /note="match: STR: Em:G50065
 match: GSS: Em:AQ014033"
 complement(58225. .58630)
 /note="match: GSS: Em:AQ140587"
 complement(58310. .58614)
 /note="match: GSS: Em:B83216"
 58331. .58743
 /note="match: GSS: Em:AQ890399"
 58927. .59200
 /note="L2 repeat: matches 1543. .1831 of consensus"
 59422. .59806
 /note="L1M4 repeat: matches 472. .862 of consensus"
 59871. .60217
 /note="L1M4 repeat: matches 867. .1199 of consensus"
 60292. .60367
 /note="L1M4 repeat: matches 1204. .1281 of consensus"
 Query Match 20.8%; Score 34.8; DB 9; Length 198917;
 Best Local Similarity 55.9%; Pred. No. 5.8;
 Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 2 ccagcgcatgtacaaagagaccgtctactccgccttcaatctgctcatgcatcattaccac 61
 Db 174349 CCAGCGCATGTACGAATATCTCTCTCATTTGAGTTCAGGTGCTGGATTACGAAA 174290
 QY 62 ccccttcggagcagggcagcaccaccccgagccgcagcccccgtggtgaagtgagcccg 119
 Db 174289 AACCTCTGCAGCTGTGTGCTGCCAACAGCTGCCATTTGTCTTGAACCCG 174232
 RESULT 15
 AC091039 AC091039 159135 bp DNA HTG 11-AUG-2001
 LOCUS Homo sapiens chromosome 18 clone RP11-403A5 map 18, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 3 unordered pieces.
 ACCESSION AC091039
 VERSION AC091039.3 GI:15148197
 KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Unpublished
 TITLE Homo sapiens chromosome 18, clone RP11-403A5
 REFERENCE 2 (bases 1 to 159135)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boduslavsky,L., Boukhallal,B., Brown,A.,
 Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
 Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Lacroque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Maheeters,R., Meldrum,J., Meneis,L.,
 McEwan,P., McKernan,K., McPheeters,R., Murphy,T., Naylor,J.,
 Mihova,T., Mlenga,V., Murphy,T., O'Donnell,P., O'Neil,D., Oliver,J.,
 Norman,C.H., O'Connot,T., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
 Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
 Theodore,J., Travers,M., Travis,N., Tridilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 11, 2001 this sequence version replaced gi:14336523.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L13167
 Center clone name: 403_A_5

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 69068: contig of 69068 bp in length
 * 69069 69168: gap of 100 bp
 * 69169 73453: contig of 4285 bp in length
 * 73454 73553: gap of 100 bp
 * 73554 159135: contig of 85582 bp in length.
 FEATURES
 Location/Qualifiers
 1. .159135
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clones="RP11-403A5"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 49874 a 28773 c 28100 g 52128 t 260 others
 ORIGIN
 Query Match 28.68; Score 34.6; DB 2; Length 159135;
 Best Local Similarity 66.2%; Pred. No. 6.8;
 Matches 49; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 16 aaagagacgcgtactcgcgttccaatctgctcatgcatcattaccacccctcgagagca 75
 Db 6618 AAAGGCACTGGTTGCACCGGTGGAGGTGCTCATGCTTACCAACACCAACCGTCA 6677
 QY 76 ggagcagcaccacca 89
 Db 6678 GGGCTGCCCTCCCA 6691

us-09-904-420a-1_copy_820_940.rge

Thu Feb 21 07:25:34 2002

Search completed: February 20, 2002, 14:24:46
Job time: 14325 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:30:45 : Search time 366.07 Seconds
(without alignments)
283.378 Million cell updates/sec

Title: US-09-904-420A-l_COPY_820_940
Perfect score: 121
Sequence: 1 accagggcatgtacaaga.....gctggttaagtggaccgcgc 121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_1101.*
1:	/SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*	2:	/SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*	
3:	/SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*	4:	/SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*	
5:	/SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*	6:	/SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*	
7:	/SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*	8:	/SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*	
9:	/SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*	10:	/SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*	
11:	/SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*	12:	/SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*	
13:	/SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*	14:	/SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*	
15:	/SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*	16:	/SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*	
17:	/SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*	18:	/SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*	
19:	/SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*	20:	/SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*	
21:	/SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*	22:	/SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
C 1	32.4	26.8	5187	AA57912	G. oxydans D-sorbi
C 2	32	26.4	2640	AAQ26664	buAT. Bos taurus.
C 3	31.2	25.8	35465	AAF54723	Nucleotide sequenc
C 4	31	25.6	6045	AAH18235	Human cDNA sequenc
C 5	30.4	25.1	6002	AAH60227	Poly(A) binding pr
C 6	30.4	25.1	9759	AAV34766	Rubella virus RA27
C 7	30.2	25.0	495	AAH74659	Nucleotide sequenc
C 8	30.2	25.0	498	AAH74658	Nucleotide sequenc
C 9	30.2	25.0	781	AAH08593	Human cDNA clone (
C 10	30.2	25.0	2540	AAH18316	Human cDNA sequenc
C 11	30.2	25.0	3177	AAT16474	B virus gB glycopr

12	30.2	25.0	3177	19	AAV33167	Simian herpesvirus
13	30.2	25.0	3250	21	AAA14621	Nucleotide sequenc
14	30.2	25.0	5635	15	AAQ68002	Ad2/CFTR-1 nucleot
15	30.2	25.0	6120	16	AAQ91273	Version of the pla
16	30.2	25.0	6420	20	AAI15625	Sequence of pGEM C
17	30.2	25.0	7507	16	AAQ91272	Plasmid vector pAd
18	30.2	25.0	7607	21	AAAS9076	Nucleotide sequenc
19	30.2	25.0	7623	18	AAAT6056	Plasmid padneo-int
20	30.2	25.0	7980	21	AAZ50627	Recombinant plasm
21	30.2	25.0	8073	21	AAQ64553	Nucleotide sequenc
22	30.2	25.0	8236	18	AAAT47202	Recombinant adenov
23	30.2	25.0	8238	21	AAAS9078	Nucleotide sequenc
24	30.2	25.0	9407	21	AAZ50628	Recombinant plasm
25	30.2	25.0	9592	17	AAAT6752	Adenovirus vector
26	30.2	25.0	9629	17	AAAT14600	pXJCL-hCMV-CSF expr
27	30.2	25.0	9641	20	AAAT15626	pGEM Age pac-beta-
28	30.2	25.0	10332	20	AAAT33921	Nucleotide sequenc
29	30.2	25.0	10398	17	AAAT15286	Ad.AV.CMVlac2 hybr
30	30.2	25.0	10398	19	AAV221130	cDNA sequence of p
31	30.2	25.0	10398	19	AAV221139	cDNA sequence of p
32	30.2	25.0	10398	19	AAV221139	cDNA sequence of p
33	30.2	25.0	10597	18	AAAT60554	Complete sequence
34	30.2	25.0	11152	19	AAV32373	Complete sequence
35	30.2	25.0	11152	21	AAAS9047	Nucleotide sequenc
36	30.2	25.0	12135	18	AAAT60555	Plasmid pAdCMVgag-
37	30.2	25.0	14455	19	AAV32374	Complete sequence
38	30.2	25.0	14455	21	AAAS9050	Nucleotide sequenc
39	30.2	25.0	31183	22	AAAD03963	Adenovirus Ad-5 de
40	30.2	25.0	31446	21	AAAO9088	AdPB-beta-galactos
41	30.2	25.0	32026	18	AAAT60559	Recombinant adenov
42	30.2	25.0	32165	21	AAAO9092	AdMTV-beta-galact
43	30.2	25.0	32165	21	AAA14723	Nucleotide sequenc
44	30.2	25.0	32166	21	AAAO9090	Ad/SA-beta-galacto
45	30.2	25.0	32166	22	AAAC89170	AdRSVphYDE region

ALIGNMENTS

RESULT 1	AA57912/c
ID	AA57912 standard; DNA; 5187 BP.
XX	AA57912;
XX	
DT	15-JUL-1999 (first entry)
XX	
DE	G. oxydans D-sorbitol dehydrogenase coding sequence.
XX	
KW	D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulononic acid; precursor;
KW	L-ascorbic acid production; ss.
XX	
OS	Gluconobacter oxydans.
XX	
PN	WO9920763-A1.
XX	
PD	29-APR-1999.
XX	
PF	13-OCT-1998; 98WO-JP04612.
XX	
PR	17-OCT-1997; 97JP-0285280.
XX	
PA	(FUJI) FUJISAWA PHARM CO LTD.
XX	
PI	Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;
XX	
DR	Poly(A) binding pr
XX	
PT	gene group for D-sorbitol dehydrogenase, useful for simple
PT	large-scale production of L-sorbose or 2-keto-L-gulononic acid as
PT	precursor for L-ascorbic acid
XX	
PS	Claim 21; Page 60-62; 83pp; Japanese.


```
XX SQ Sequence 6002 BP; 1405 A; 1340 C; 1697 G; 1559 T; 1 other;
Query Match 25.1%; Score 30.4; DB 20; Length 6002;
Best Local Similarity 53.3%; Pred. No. 8.2;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 2 ccagcgcatgtacaaagagacgtctactcccgcttcaatctgctcatgctacccac 61
Db 1469 CCAGCAGCAGCTCTCAGGCTCCAGCTCCAGACTCCAGCGCGTTCGCTACTCCCG 1410
QY 62 cccctcgggagcaggcagcaccgccagccgagcccccgcgtggttaagtgcacccgcc 121
Db 1409 CGCCCGCGGGCGCCCTCCCGGCTCCCGACCGCGCGGCGACAGAAGATGGCGCGCC 1350
RESULT 6
AAV34766
ID AAV34766 standard; RNA: 9759 BP.
AC AAV34766;
DT 11-SEP-1998 (first entry)
DE Rubella virus RA27/3 genomic sequence.
XX genomic sequence; viral genome; rubella; RA27/3; vaccine; CRS;
KW congenital rubella syndrome; vaccination; ss.
XX Rubella virus.
XX Key Location/Qualifiers
FH 41..6388
FT CDS /tag= a
FT /product= NSP-ORF, nonstructural proteins
FT 6509..9700
FT CDS /tag= b
FT /product= SP-ORF, virion structural proteins C,E2,E1
XX WO9820901-A1.
XX 22-MAY-1998.
XX 07-NOV-1997; 97WO-US20399.
XX 12-NOV-1996; 96US-0030734.
XX (UYGE-) UNIV GEORGIA STATE RES FOUND.
XX Frey TK, Pougatchev KV;
XX WPI: 1998-297616/26.
XX P-PSDB: AAW59276, AAW59277.
XX Genomic sequence for RA27/3 strain of rubella virus - useful to
XX identify further attenuated strains of rubella virus, e.g. for use
XX in vaccines for congenital rubella syndrome
XX Claim 3; Fig 1; 34pp; English.
XX The sequence is that of the rubella RA27/3 genome.
XX The nucleic acid can be used to identify further attenuated
XX strains of rubella virus, by sequencing putative strains and determining
XX whether specific attenuating mutation sites are present by comparison
XX with the sequence of RA27/3. Such strains, and the nucleic acid
XX sequence, are useful in vaccines against rubella virus, an
XX important human pathogen. Whilst acute infection is usually benign,
XX infection during early pregnancy can result in viral passage across the
XX placenta and replication in the foetus, causing severe birth defects
XX (congenital rubella syndrome, CRS). Eight live attenuated rubella virus
XX vaccines to prevent CRS have previously been developed by serial passage
XX through culture, and vaccine comprising the RA27/3 strain is the most
```

```
CC commonly used. Genetic characterisation of RA27/3 is also important to
CC assure that the sequence (and therefore the attenuated phenotype) is
CC maintained during vaccine production, and to determine whether
CC complications sometimes occurring after vaccination (e.g. chronic
CC arthritis) are of wild-type or vaccine origin. The nucleic acid may
CC also be used to identify specific attenuating mutation sites, by
CC introducing RA27/3-specific nucleotides into a wild-type infectious
CC clone (e.g. an existing infectious clone based on the Therien strain)
CC and determining whether the resulting virus is attenuated. It may
CC also be used to express RA27/3 in recombinant host cells, by
CC transfecting host cells with the expression vector containing
CC the nucleic acid and culturing the cells under suitable conditions.
XX SQ Sequence 9759 BP; 1458 A; 3775 C; 3008 G; 1518 T; 0 other;
Query Match 25.1%; Score 30.4; DB 19; Length 9759;
Best Local Similarity 57.3%; Pred. No. 9;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 25 gtctactcgcgttcaatctgctcatgcatlaccaccccccctcggagcaggcgacac 84
Db 5986 gtggacccccgcgaggtgggtcttctcggttccacatcccggtgaagcatgtgagcac 6045
QY 85 cccacagcgcgacccccgcgtgtaagtggacccgc 120
Db 6046 ccctacccccagcttctcgggcacgtcgccaccgc 6081
RESULT 7
AAH74659
ID AAH74659 standard; DNA: 495 BP.
AC AAH74659;
DT 15-OCT-2001 (first entry)
DE Nucleotide sequence of C-terminal fusion protein of pIX protein.
XX pIX protein; chimeric protein; gene transfer vector; ss.
XX Synthetic.
OS Mastadenovirus.
XX Key Location/Qualifiers
FH 1..498
FT CDS /tag= a
FT /product= "chimeric protein"
XX WO200158940-A2.
XX 16-AUG-2001.
XX 09-FEB-2001; 2001WO-US04233.
XX 09-FEB-2000; 2000US-0181163.
XX (GENV-) GENVEC INC.
XX Roelvink PW, Kovcsdi I, Wickham TJ;
XX WPI: 2001-497066/54.
XX P-PSDB: AAG63606.
XX Chimeric pIX protein useful in an adenovirus gene transfer vector for
XX infecting cells comprises at least one adenoviral pIX domain and a
XX non-native amino acid .
XX Example 1; Page 27; 28pp; English.
XX The present sequence encodes a chimeric fusion protein of the invention.
XX The specification describes a chimeric pIX protein having at least one
XX adenoviral pIX domain and a non-native amino acid sequence. The
```


XX
SQ Sequence 781 BP; 231 A; 157 C; 236 G; 154 T; 3 other;

Query Match 25.0%; Score 30.2; DB 22; Length 781;
Best Local Similarity 69.5%; Pred. No. 6.6;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 61 ccccccctggagcaggcgacgaccgccgcagccccccctgtgaatggacccg 119
|| | ||| ||||||| || | || || || || || || || || || || ||
Db 49 ccgcgagccaccagggcagcagccgcgacccgcgcgcctggccaaggagccg 107

RESULT 10
AAH18316
ID AAH18316 standard; cDNA: 2540 BP.
XX
AC AAH18316;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18318.
XX
DE Human: primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
KW Homo sapiens.
OS
XX EP1074617-A2.
PN
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 93JP-0248036.
PR 27-AUG-1999; 93JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
LR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
XX
PS Claim 8; SEQ ID 18318; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
of the present invention.

XX
SQ Sequence 2540 BP; 790 A; 519 C; 637 G; 594 T; 0 other;

Query Match 25.0%; Score 30.2; DB 22; Length 2540;
Best Local Similarity 69.5%; Pred. No. 8.1;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 61 cccccctggagcaggcgacgaccgccgcagccccccctgtgaatggacccg 119
|| | ||| ||||||| || | || || || || || || || || || ||
Db 49 ccgcgagccaccagggcagcagccgcgacccgcgcgcctggccaaggagccg 107

RESULT 11
AAT16474
ID AAT16474 standard; DNA: 3177 BP.
XX
AC AAT16474;
XX
DT 11-MAY-1996 (first entry)
XX
DE B virus gB glycoprotein and ICP 18.5 kDa protein coding sequence.
XX
KW Herpes simian monkey B virus gB glycoprotein; UL27; diagnosis;
herpes B virus; ICP 18.5 kDa protein; UL28; ss.
XX
OS Herpes simian monkey B virus.
XX
FH Key Location/Qualifiers
CDS 1..249
FT /*tag= a
/*product= ICP_18.5_kDa_protein_fragment
269..2944
FT CDS /*tag= b
/*product= gB_glycoprotein
FT US5487969-A.
PN
XX
XX 30-JAN-1996.
XX
XX 01-APR-1993; 93US-0042747.
XX
XX 01-APR-1993; 93US-0042747.
XX (SWBI-) SOUTHWEST FOUND BIOMEDICAL, RES.
XX
PI Black D, Eberle R, Hilliard J, Scinicariello F;
XX
XX WPI: 1996-105220/11.
DR P-PSDB; AAR92745, AAR92746.
XX
XX Detection of herpes B virus by PCR amplification of sample DNA - to
detect a specific herpes simian monkey B virus DNA segment.
XX
XX Claim 1; Column 11-18; 22pp; English.

CC This sequence encodes the herpes simian monkey B virus gB
glycoprotein (UL27) and a portion of an ICP 18.5 kDa protein (UL28),
as well as some non-coding sequences. The sequence has been used as
a basis for the development of differential diagnostic tests for
rapid testing of suspect B virus cases in humans and monkeys.
CC Comparison of the B virus gB proteins-glycoprotein sequence with
sequences of gBs of other primate alpha-herpes viruses allows
identification of regions which are strongly conserved among all
primate alpha-herpes viruses and other regions which are highly
divergent among them.

XX
SQ Sequence 3177 BP; 494 A; 1187 C; 1033 G; 463 T; 0 other;

Query Match 25.0%; Score 30.2; DB 17; Length 3177;

Best Local Similarity 53.9%; Pred. No. 8.4; Mismatches 62; Conservative 0; Gaps 0;

QY 6 cggcatgtacaagagacgctctactccgcttcaatctctcatctcattaccaccccc 65
 150 cgtggtgtacgacgcgagctctctcattctctactcgtctcgtcagacactgcccc 209

Db 150 cgtggtgtacgacgcgagctctctcattctctactcgtctcgtcagacactgcccc 209

QY 66 ctgggagcagggcagcccccagccgagcccccgtgtgaagtggacccgc 120
 210 caggctcgcgcggggggccgacagcccccgtagccgcccgcgcgcgc 264

Db 210 caggctcgcgcggggggccgacagcccccgtagccgcccgcgcgcgc 264

RESULT 12
 AAV33167
 ID AAV33167 standard; DNA; 3177 BP.
 XX
 AC AAV33167;
 XX
 DT 06-NOV-1998 (first entry)
 XX
 DE Simian herpesvirus B DNA sequence coding for a gB glycoprotein.
 XX
 KW Simian herpesvirus B gB glycoprotein; UL27; ICP protein; UL28;
 KW differential diagnostic test; immunoassay; antibody; ss.
 XX
 OS Simian herpesvirus B.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..249
 FT /tag= a
 FT /note= "encodes a fragment of an ICP Protein (UL28)"
 FT CDS 269..2944
 FT /tag= b
 FT /product= "gB glycoprotein (UL27)"
 XX
 US5767265-A.
 XX
 16-JUN-1998.
 XX
 10-OCT-1995; 95US-0541878.
 XX
 01-APR-1993; 93US-0042747.
 PR 10-OCT-1995; 95US-0541878.
 XX
 (SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.
 XX
 PI Black D, Eberle R, Hilliard J, Scinicariello F;
 DR WPI; 1998-361791/31.
 DR P-PSDB; AAW70293, AAW70294.
 XX
 Monkey herpes B virus DNA - coding for gB glycoproteins and polypeptides
 PT
 Claim 1; Columns 11-18; 22pp; English.
 XX
 The present sequence represents the Simian herpesvirus B DNA sequence coding for a gB glycoprotein (UL27; AAW70293) and a portion of an ICP 18.5 kDa protein (UL28; AAW70294). The invention uses these DNA and protein sequences as a basis for the development of differential diagnostic tests for the rapid identification of Simian herpesvirus B cases. Therefore, the virus can be detected by detecting the DNA sequence and knowledge of the amino acid sequence will help in the design of DNA probes and of peptides for use in immunoassays and for antibody production.
 CC
 Sequence 3177 BP; 494 A; 1187 C; 1033 G; 463 T; 0 other;
 XX

Query Match 25.0%; Score 30.2; DB 19; Length 3177;
 Best Local Similarity 53.9%; Pred. No. 8.4; Mismatches 62; Conservative 0; Gaps 0;

QY 6 cggcatgtacaagagacgctctactccgcttcaatctctcatctcattaccaccccc 65
 150 cgtggtgtacgacgcgagctctctcattctctactcgtctcgtcagacactgcccc 209

Db 150 cgtggtgtacgacgcgagctctctcattctctactcgtctcgtcagacactgcccc 209

QY 66 ctgggagcagggcagcccccagccgagcccccgtgtgaagtggacccgc 120
 210 caggctcgcgcggggggccgacagcccccgtagccgcccgcgcgcgc 264

Db 210 caggctcgcgcggggggccgacagcccccgtagccgcccgcgcgcgc 264

RESULT 13
 AAA14621
 ID AAA14621 standard; DNA; 3250 BP.
 XX
 AC AAA14621;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Nucleotide sequence of a recombinant rabies andenovirus construct.
 XX
 KW Adenovirus; Immune response; rabies; E1 region; glycoprotein; vaccine;
 KW intron; ss.
 XX
 OS Synthetic.
 OS Mastadenovirus.
 XX
 FH Key Location/Qualifiers
 FT promoter 359..700
 FT /tag= a
 FT /note= "HCMV promoter"
 FT intron 798..999
 FT /tag= b
 FT misc_feature 1051..2625
 FT /tag= c
 FT /note= "DNA encoding rabies glycoprotein"
 FT polyA_signal 2662..2800
 FT /tag= d
 FT /note= "SV40 polyA sequences"
 XX
 CA2246626-AL.
 XX
 08-APR-1999.
 XX
 08-OCT-1998; 98CA-2246626.
 XX
 08-OCT-1997; 97US-0070452.
 XX
 (MICR-) MICROBIX BIOSYSTEMS INC.
 XX
 Hughes K, Cassidy K, Yarosh O, Sparkuhl Y, Graham FL, Moore M;
 PI Prevec L;
 XX
 WPI; 2000-318453/28.
 XX
 Recombinant adenovirus expressing rabies glycoprotein, useful for vaccination against rabies, particularly of wild animals -
 PT
 Example 3; Fig 9a-d; 54pp; English.
 XX
 The specification describes a recombinant adenovirus that raises an immune response against rabies. The adenovirus has a functionally disrupted E1 region and includes an expression cassette containing DNA that encodes rabies glycoprotein, linked to control sequences functional in eukaryotic cells. The adenovirus is used in vaccines against rabies, particularly oral vaccines for protecting wild animals that are vectors of the disease. Incorporation of an intron into the CC adenovirus prevents homologous recombination with E1 regions endogenous to host cells, i.e. to prevent development of replication competent CC versions of the adenovirus. The present sequence represents a CC recombinant rabies adenovirus construct AdBHG3E3CA13(1)RG.
 XX
 Sequence 3250 BP; 810 A; 740 C; 866 G; 834 T; 0 other;
 SQ

```

Query Match      25.0%; Score 30.2; DB 21; Length 3250;
Best Local Similarity 58.2%; Pred. No. 8.4;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 30 ctccgccttcaatctgcattgcattaccacccccctcggagcagggcagcaccacca 89
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3116 cgcgggattgtgactgacttcttctgagcccgcttgcacagcagtgacgttcccg 3175

QY 90 gccgcagccccgcgtgtaagtgcacccgc 120
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3176 tcattccgcgcgcatgacaaagttgacggctc 3206

RESULT 14
ID AAQ68002 standard; DNA; 5635 BP.
XX
AC AAQ68002;
XX
DT 26-OCT-1995 (first entry)
DE
XX
XX Ad2/CFTR-1 nucleotide sequence.
XX Recombinant adenovirus; Ad2/CFTR-1; adenovirus 2 serotype; Ela; Elb;
KW viral replication; gene expression; gene therapy; cystic fibrosis;
KW cystic fibrosis transmembrane conductance regulator; CFTR;
KW promoter; E3; p19; MHC; class I; viral latency; pulmonary airway; ds.
XX
OS Homo sapiens.
OS Adenoviral sp.
XX
FH Key
FH repeat_region 1..104
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Represents the origin of replication"
FT enhancer 190..380
FT /*tag= b
FT /function= E1A enhancer and viral packaging domain
FT promoter 380..500
FT /*tag= C
FT /note= "E1A promoter region"
FT prim_transcript 499..5635
FT /*tag= d
FT /note= "Hybrid E1A-CFTR-E1B message"
FT 5'UTR 499..546
FT /*tag= e
FT misc_feature 547..595
FT /*tag= f
FT /note= "Synthetic linker sequences"
FT misc_feature 593..5093
FT /*tag= g
FT /note= "Represents nucleotides 123-4622 of the published
FT CFTR cDNA sequence"
FT CDS 603..5045
FT /*tag= h
FT /product= CFTR
FT 3'UTR 5093..5635
FT /*tag= i
FT /note= "E1B 3' UTR"
FT intron 5099..5190
FT /*tag= j
FT /note= "E1B 3' intron"
FT prim_transcript 5177..5635
FT /*tag= k
FT /note= "IX protein mRNA"
FT CDS 5201..5623
FT /*tag= l
FT /product= IX protein (hexon-associated protein)
XX
XX WC9412649-A.
XX
XX 09-JUN-1994.
PN
PD

```

```

XX PF 02-DEC-1993; 93WO-US11667.
XX PR 03-DEC-1992; 92US-0985478.
XX PR 01-OCT-1993; 93US-0130682.
XX PR 13-OCT-1993; 93US-0136742.
XX (GENZ ) GENZYME CORP.
XX
XX Armentano D, Couture LA, Gregory RJ, Smith AE;
XX WPI: 1994-200277/24.
XX P-PSDB; AAR79011-12.
XX
XX Adeno:virus-based gene therapy vectors - esp. useful for gene
XX therapy of cystic fibrosis
XX
XX Claim 4; Page 67-80; 167pp; English.
XX
XX This sequence represents the nucleotide sequence of the recombinant
XX adenovirus Ad2/CFTR-1. This virus is derived from the relatively
XX benign adenovirus 2 serotype. The Ela and E1b regions of the viral
XX genome, which are involved in the early stages of viral replication
XX have been deleted which impairs viral gene expression and viral
XX replication. The cystic fibrosis transmembrane conductance
XX regulator (CFTR) coding sequence is inserted into the genome in
XX place of the Ela/E1b region and transcription of the CFTR sequence
XX is driven by the endogenous Ela promoter. This is a moderately
XX strong promoter that is functional in a variety of cells. This
XX adenovirus retains the E3 viral coding region. As a consequence the
XX length of the adenovirus-CFTR DNA is greater than that of wild type
XX adenovirus. This renders the DNA more difficult to package and means
XX that the growth of the Ad2/CFTR virus is impaired even in permissive
XX cells that provide the missing Ela and E1b functions. The E3 region
XX encodes a number of proteins, including p19 which is believed to
XX interact with and prevent presentation of MHC class I proteins. This
XX property prevents recognition of the infected cells and thus may allow
XX viral latency. This adenovirus may be administered to the pulmonary
XX airways in the gene therapy of cystic fibrosis.
XX
XX Sequence 5635 BP; 1619 A; 1142 C; 1324 G; 1550 T; 0 other;

```

```

Query Match      25.0%; Score 30.2; DB 15; Length 5635;
Best Local Similarity 58.2%; Pred. No. 9.3;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```

```

QY 30 ctccgccttcaatctgcattgcattaccacccccctcggagcagggcagcaccacca 89
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5417 cgcgggattgtgactgacttcttctgagcccgcttgcacagcagtgacgttcccg 5476

QY 90 gccgcagccccgcgtgtaagtgcacccgc 120
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5477 tcattccgcgcgcatgacaaagttgacggctc 5507

```

```

RESULT 15
AAQ91273
ID AAQ91273 standard; DNA; 6120 BP.
XX
XX AAQ91273;
XX
XX 26-NOV-1995 (first entry)
XX
XX Version of the plasmid vector pAdCMV-HS-vector.
XX
XX Adenovirus-5 gene expression system; Ad5; cytomegalovirus promoter;
XX mouse beta-globin; polyadenylation signal; ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX rep_origin 1..353

```

Thu Feb 21 07:25:35 2002

```
FT      /*tag= a
FT      /label= Ad 5 ori
PN      W09516772-A.
XX
XX      22-JUN-1995.
PD
XX
XX      14-DEC-1994; 94WO-US14502.
PF
XX
XX      14-DEC-1993; 93US-0166925.
PR
XX
XX      (CORR ) CORNELL RES FOUND INC.
PA
XX      Falckpedersen ES;
PI
XX
XX      WPI; 1995-231566/30.
DR
XX
XX      New adenovirus-5 gene expression system - used for highly efficient,
PT      recombinant expression of heterologous genes or gene prods. in host
PT      cells
XX
XX      Disclosure; Fig 23; 86pp; English.
PS
XX
XX      Plasmid vector pADCMW-HS-vector comprises the following elements.
CC      Adenovirus nt sequence from 1-353 contg. origin of replication and
CC      the viral packaging sequence; Not I site of insertion for second
CC      gene; cytomegalovirus (CMV) early promoter/enhancer; splice donor
CC      and splice acceptor; Hind III site; stuffer site for cDNA insertion;
CC      Sal I site; mouse beta-globin polyadenylation sequence; Ad 5 genome
CC      sequence from 354-2800 for overlap recombination; plasmid vector
CC      sequence. The plicing elements are from pMI-1S cat. The plasmid
CC      vector is converted into a recombinant adenovirus for expression of
CC      a heterologous gene(s) and/or gene product(s) in a mammalian cell.
XX
XX      Sequence 6120 BP; 1328 A; 1533 C; 1721 G; 1538 T; 0 other;
SQ

Query Match      25.0%; Score 30.2; DB 16; Length 6120;
Best Local Similarity 58.2%; Pred. No. 9.4;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Oy      30 ctcgccttcaatctgctcatgcattaccacccccctcgaggagcgagcgacacccca 89
Db      2516 cgcgggattgtgacgactttgcttcttcctgagcccgctgcaagcaytgagcttcccg 2575

Oy      90 gccgcagccccgcctggtaagtggacccgc 120
Db      2576 tcattccgcgcgcatgacaagtgtgacggctc 2606

Search completed: February 20, 2002, 14:30:58
Job time: 12447 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:32:12 ; Search time 162.84 Seconds
(without alignments)
168.287 Million cell updates/sec

Title: US-09-904-420A-1_COPY_820_940
Perfect score: 121
Sequence: 1 accagggcatgtacaaga.....gctggtaagtgagacccgcc 121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30.2	25.0	1796	1	US-08-540-077-2
2	30.2	25.0	1800	1	US-08-540-077-4
3	30.2	25.0	3177	1	US-08-042-747A-4
4	30.2	25.0	5635	1	US-08-136-742A-3
5	30.2	25.0	5635	3	US-09-248-026-3
6	30.2	25.0	5635	5	PCT-US93-11667-3
7	30.2	25.0	6420	2	US-08-374-483-1
8	30.2	25.0	8236	1	US-08-461-837-1
9	30.2	25.0	8236	3	US-08-973-223-1
10	30.2	25.0	8236	4	US-09-347-060-1
11	30.2	25.0	8236	5	PCT-US96-09495-1
12	30.2	25.0	9592	1	US-08-393-734-3
13	30.2	25.0	9592	4	US-08-894-489-3
14	30.2	25.0	9641	2	US-08-374-483-3
15	30.2	25.0	10398	2	US-08-331-384-1
16	30.2	25.0	10398	2	US-08-708-188-1
17	30.2	25.0	10398	2	US-08-836-087-1
18	30.2	25.0	10398	4	US-09-246-320-1
19	30.2	25.0	10398	4	US-09-242-743-1
20	30.2	25.0	34303	2	US-08-735-609-4
21	30.2	25.0	34303	2	US-08-735-609-4
22	30.2	25.0	34303	3	US-09-315-372-4
23	30.2	25.0	34303	3	US-09-244-752-4
24	30.2	25.0	34303	3	US-09-245-497-4
25	30.2	25.0	34382	2	US-08-374-483-6
26	30.2	25.0	35408	4	US-08-973-334-3
27	30.2	25.0	35408	4	US-09-563-869A-3

28	30.2	25.0	35408	4	US-08-549-489-3
29	30.2	25.0	35935	2	US-08-735-609-1
30	30.2	25.0	35935	2	US-08-735-609-1
31	30.2	25.0	35935	3	US-08-379-452-43
32	30.2	25.0	35935	3	US-09-315-372-1
33	30.2	25.0	35935	3	US-09-244-752-1
34	30.2	25.0	35935	3	US-09-245-497-1
35	30.2	25.0	35935	3	US-09-409-670-43
36	29.8	24.6	2308	4	US-09-382-256-9
37	29.8	24.6	2308	4	US-09-395-115-9
38	29.4	24.3	9757	1	US-08-093-453B-1
39	29.4	24.3	9759	1	US-08-459-041A-1
40	29.4	24.3	9759	3	US-08-999-733-1
41	27.8	23.0	3777	3	US-09-121-321-15
42	27.8	23.0	3777	4	US-08-933-803A-15
43	27.6	22.8	4533	4	US-09-171-410-2
44	27.4	22.6	1483	4	US-09-262-749-1
45	27.4	22.6	1553	3	US-09-022-669-1

ALIGNMENTS

RESULT 1
US-08-540-077-2
; Sequence 2, Application US/08540077
; Patent No. 5824544
; GENERAL INFORMATION:
; APPLICANT: ARMENTANO, DONNA
; APPLICANT: ROMANCZUK, HELEN
; APPLICANT: WADSWORTH, SAMUEL C.
; TITLE OF INVENTION: NOVEL ADENOVIRUS VECTORS
; TITLE OF INVENTION: FOR GENE THERAPY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540.077
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/409.874
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Seide, Rochelle K
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: A30086A - 2415/31081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2626
; TELEFAX: 212-765-2519
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-540-077-2

Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 43, Appl
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 43, Appl
Sequence 9, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 15, Appl
Sequence 15, Appl
Sequence 2, Appli
Sequence 1, Appli
Sequence 1, Appli

Db	778	CGCGGGATTGTGACTGACTTTGCTTTCTGTGAGCCCGCTTGCACAGCAGTGCAGCTTCCCGT	837
Qy	90	gcgcagcccccgctggttaagtggacccgc	120
Db	838	TCATCCGCCCGGATGACAAAGTTGACGGCTC	868
RESULT			
US-08-042-747A-4			
: Sequence 4, Application US/08042747A			
: Patent No. 5487969			
: GENERAL INFORMATION:			
: APPLICANT: Eberle, Richard			
: APPLICANT: Black, Darla			
: APPLICANT: Scinicariello, Franco			
: APPLICANT: Hilliard, Julia K.			
: TITLE OF INVENTION: Cloning and Amplification of Monkey B			
: TITLE OF INVENTION: Virus Genes			
: NUMBER OF SEQUENCES: 12			
: CORRESPONDENCE ADDRESS:			
: ADDRESSEE: Cox & Smith Incorporated			
: STREET: 112 East Pecan Street, Suite 2000			
: CITY: San Antonio			
: STATE: Texas			
: COUNTRY: USA			
: ZIP: 78205			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: Floppy disk			
: COMPUTER: IBM PC compatible			
: OPERATING SYSTEM: PC-DOS/MS-DOS			
: SOFTWARE: PatentIn Release #1.0, Version #1.25			
: CURRENT APPLICATION DATA:			
: FILING DATE:			
: APPLICATION NUMBER: US/08/042.747A			
: CLASSIFICATION: 435			
: ATTORNEY/AGENT INFORMATION:			
: NAME: Haymond, W. Bradley			
: REGISTRATION NUMBER: 35186			
: REFERENCE/DOCKET NUMBER: S-0072.179			
: TELECOMMUNICATION INFORMATION:			
: TELEPHONE: 210-554-5500			
: TELEFAX: 210-226-8395			
: TELEX: 767609			
: INFORMATION FOR SEQ ID NO: 4:			
: SEQUENCE CHARACTERISTICS:			
: LENGTH: 3177 base pairs			
: TYPE: nucleic acid			
: STRANDEDNESS: single			
: TOPOLOGY: linear			
: MOLECULE TYPE: Genomic DNA			
: FEATURE:			
: NAME/KEY: CDS			
: LOCATION: 1..2941			
: FEATURE:			
: NAME/KEY: CDS			
: LOCATION: 1..249			
US-08-042-747A-4			
Query Match 25.0%; Score 30.2; DB 1; Length 3177;			
Best Local Similarity 53.9%; Pred. No. 2.9;			
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;			
Qy	6	cggcatgtacaaagagagaccccgcttactccgccttcattatctgtcatgcaatccaccacccc	65
Db	150	CGTGGTGTACACCGCGACGTCTTCTCCATCTCTACTCGGTGCTGACAGCACCTGGGCCCC	209
Qy	66	ctcgggagcagggcagcaccacccagcgcagcccccgctggtaagtgacccccgc	120
Db	210	CAGGCTCGCGCGCGGGGGCCGACGACCGCCCGGTAGCGCGCGCGCGCGG	264
RESULT 4			

Db	778	CGCGGGATTGTGACTGACTTTGCTTTCTGTGAGCCCGCTTGCACAGCAGTGCAGCTTCCCGT	837
Qy	90	gcgcagcccccgctggttaagtggacccgc	120
Db	838	TCATCCGCCCGGATGACAAAGTTGACGGCTC	868
RESULT			
US-08-042-747A-4			
: Sequence 4, Application US/08042747A			
: Patent No. 5487969			
: GENERAL INFORMATION:			
: APPLICANT: Eberle, Richard			
: APPLICANT: Black, Darla			
: APPLICANT: Scinicariello, Franco			
: APPLICANT: Hilliard, Julia K.			
: TITLE OF INVENTION: Cloning and Amplification of Monkey B			
: TITLE OF INVENTION: Virus Genes			
: NUMBER OF SEQUENCES: 12			
: CORRESPONDENCE ADDRESS:			
: ADDRESSEE: Cox & Smith Incorporated			
: STREET: 112 East Pecan Street, Suite 2000			
: CITY: San Antonio			
: STATE: Texas			
: COUNTRY: USA			
: ZIP: 78205			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: Floppy disk			
: COMPUTER: IBM PC compatible			
: OPERATING SYSTEM: PC-DOS/MS-DOS			
: SOFTWARE: PatentIn Release #1.0, Version #1.25			
: CURRENT APPLICATION DATA:			
: FILING DATE:			
: APPLICATION NUMBER: US/08/042.747A			
: CLASSIFICATION: 435			
: ATTORNEY/AGENT INFORMATION:			
: NAME: Haymond, W. Bradley			
: REGISTRATION NUMBER: 35186			
: REFERENCE/DOCKET NUMBER: S-0072.179			
: TELECOMMUNICATION INFORMATION:			
: TELEPHONE: 210-554-5500			
: TELEFAX: 210-226-8395			
: TELEX: 767609			
: INFORMATION FOR SEQ ID NO: 4:			
: SEQUENCE CHARACTERISTICS:			
: LENGTH: 3177 base pairs			
: TYPE: nucleic acid			
: STRANDEDNESS: single			
: TOPOLOGY: linear			
: MOLECULE TYPE: Genomic DNA			
: FEATURE:			
: NAME/KEY: CDS			
: LOCATION: 1..2941			
: FEATURE:			
: NAME/KEY: CDS			
: LOCATION: 1..249			
US-08-042-747A-4			
Query Match 25.0%; Score 30.2; DB 1; Length 3177;			
Best Local Similarity 53.9%; Pred. No. 2.9;			
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;			
Qy	6	cggcatgtacaaagagagaccccgcttactccgccttcattatctgtcatgcaatccaccacccc	65
Db	150	CGTGGTGTACACCGCGACGTCTTCTCCATCTCTACTCGGTGCTGACAGCACCTGGGCCCC	209
Qy	66	ctcgggagcagggcagcaccacccagcgcagcccccgctggtaagtgacccccgc	120
Db	210	CAGGCTCGCGCGCGGGGGCCGACGACCGCCCGGTAGCGCGCGCGCGCGG	264
RESULT 4			

```

US-08-136-742A-3
; Sequence 3, Application US/08136742A
; Patent No. 5670488
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.A., Smith,
; APPLICANT: A.E.
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE & RAYMOND
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09248.026
; FILING DATE: 10-FEB-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; FILING DATE: 16-JUL-1997
; APPLICATION NUMBER: US 08/895,194
; NAME: Seide, Rochelle K.
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: A30668-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 705-5000
; TELEFAX: (212) 705-5020
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-248-026-3

Query Match 25.0%; Score 30.2; DB 3; Length 5635;
Best Local Similarity 58.2%; Pred. No. 3.3;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps

QY 30 ctcgcctcaatctgctatgctaccacccccctcgaggagcagcagccccc 89
Db 5417 CGCGGGATTGTGACTGACTTTCTTCTGACCGCGCTTGCACAGCTTCCCGT 5476
QY 90 gccgcagccccgcctgtaagtggacccgc 120
Db 5477 TCATCCGCCGGCATGACAAAGTTGACGGCTC 5507

RESULT 5
PCT-US93-11667-3
; Sequence 3, Application PC/TUS9311667
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.A., Smith,
; APPLICANT: A.E.
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11667
; FILING DATE: 02-DEC-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,478
; FILING DATE: 02-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanley, Elizabeth A.
; REGISTRATION NUMBER: 33,505

```



```
; REFERENCE/DOCKET NUMBER: NZ1-014CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
PCT-US93-11667-3

Query Match 25.0%; Score 30.2; DB 5; Length 5635;
Best Local Similarity 58.2%; Pred. No. 3.3;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 30 ctccgcttcaatctctcatcattaccacccccctcgaggagcagcagccccc 89
Db 5417 CGCGGATTTGTGACTGACTTTCTTCTGAGCCGCTTGCAAGCAGTTCCTCCGT 5476

Qy 90 gccgcagcccccgctgtaagtggaccccg 120
Db 5477 TCATCCGCCCGCGATGACAAAGTTGACGGCTC 5507

RESULT 7
US-08-374-483-1
; Sequence 1, Application US/08374483
; Patent No. 5860102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-1

Query Match 25.0%; Score 30.2; DB 2; Length 6420;
Best Local Similarity 58.2%; Pred. No. 3.4;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
Qy 30 ctccgcttcaatctctcatcattaccacccccctcgaggagcagcagccccc 89
Db 2322 CGCGGATTTGTGACTGACTTTCTTCTGAGCCGCTTGCAAGCAGTTCCTCCGT 2381

Qy 90 gccgcagcccccgctgtaagtggaccccg 120
Db 2382 TCATCCGCCCGCGATGACAAAGTTGACGGCTC 2412

RESULT 8
US-08-461-837-1
; Sequence 1, Application US/08461837
; Patent No. 5698202
; GENERAL INFORMATION:
; APPLICANT: Ertl, Hildegund C.J.
; APPLICANT: Wilson, James M.
; TITLE OF INVENTION: A Replication-Defective Adenovirus Human
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,837
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: UPNH1290USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8236 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1185..2756
US-08-461-837-1

Query Match 25.0%; Score 30.2; DB 1; Length 8236;
Best Local Similarity 58.2%; Pred. No. 3.5;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 30 ctccgcttcaatctctcatcattaccacccccctcgaggagcagcagccccc 89
Db 3557 CGCGGATTTGTGACTGACTTTCTTCTGAGCCGCTTGCAAGCAGTTCCTCCGT 3616

Qy 90 gccgcagcccccgctgtaagtggaccccg 120
Db 3617 TCATCCGCCCGCGATGACAAAGTTGACGGCTC 3647

RESULT 9
US-08-973-223-1
; Sequence 1, Application US/08973223
; Patent No. 6019978
```



```

; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9592 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-894-489-3

Query Match 25.0%; Score 30.2; DB 4; Length 9592;
Best Local Similarity 58.2%; Pred. No. 3.6;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 30 ctccgccttcaatctgctcattaccacccccctggagcagggcagcaccacca 89
Db 4913 CCGGGATTGTGACTGCTTTCTTCCCTGAGCCCGCTTGCACGAGTCCCGT 4972

QY 90 gccgcagccccgcctgtaagtggaccgcgc 120
Db 4973 TCATCCGCCCGCATGACAAGTTGACGGCTC 5003

RESULT 14
US-08-374-483-3
; Sequence 3, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4100
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-3

Query Match 25.0%; Score 30.2; DB 2; Length 9641;
Best Local Similarity 58.2%; Pred. No. 3.6;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 30 ctccgccttcaatctgctcattaccacccccctggagcagggcagcaccacca 89

; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9592 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-894-489-3

Query Match 25.0%; Score 30.2; DB 4; Length 9592;
Best Local Similarity 58.2%; Pred. No. 3.6;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 30 ctccgccttcaatctgctcattaccacccccctggagcagggcagcaccacca 89
Db 4913 CCGGGATTGTGACTGCTTTCTTCCCTGAGCCCGCTTGCACGAGTCCCGT 4972

QY 90 gccgcagccccgcctgtaagtggaccgcgc 120
Db 4973 TCATCCGCCCGCATGACAAGTTGACGGCTC 5003

RESULT 15
US-08-331-384-1
; Sequence 1, Application US/08331384
; Patent No. 5856152
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and Methods
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,384
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: UPNG11490SA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10398 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-331-384-1

Query Match 25.0%; Score 30.2; DB 2; Length 10398;
Best Local Similarity 58.2%; Pred. No. 3.7;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 30 ctccgccttcaatctgctcattaccacccccctggagcagggcagcaccacca 89
Db 5719 CGCGGATTGTGACTGCTTTCTTCCCTGAGCCCGCTTGCACGAGTCCCGT 5778

QY 90 gccgcagccccgcctgtaagtggaccgcgc 120
Db 5779 TCATCCGCCCGCATGACAAGTTGACGGCTC 5809

Search completed: February 20, 2002, 14:32:27
Job time: 14406 sec
```

us-09-904-420a-1_copy_820_940.rni

Thu Feb 21 07:25:36 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 13:52:27 : Search time 3793.25 Seconds
(without alignments)
342.777 Million cell updates/sec

Title: US-09-904-420A-1_COPY_820_940
Perfect score: 121
Sequence: 1 accagcgcatgtacaaaga.....gctgtaagtggaccgcgc 121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estim:*
4: em_estom:*
5: em_estpi:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_Other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	103	85.1	273	10	AA311787	AA311787 EST182503
2	88.4	73.1	353	11	BG386035	BG386035 602455204
3	88	72.7	400	11	BG655601	BG655601 ia84a02.x
4	85.8	70.9	649	11	B1154476	B1154476 602904810
5	76.8	63.5	264	10	BB584341	BB584341 BB584341
6	66	54.5	201	11	BF799331	BF799331 RC6-C1012
C 7	62.8	51.9	288	10	A1049127	A1049127 ub35d04.r
C 8	62.8	51.9	330	10	AA958693	AA958693 vz94h11.r
9	52.2	43.1	152	10	A1131784	A1131784 uc37a06.r
10	34.8	28.8	465	13	A0137897	A0137897 HS_3058_B
11	33.8	27.9	587	11	BF538482	BF538482 602053985
C 12	32.8	27.1	411	11	BF846035	BF846035 PM1-EN006

C 13	32.6	26.9	192	10	AV070409	AV070409
C 14	32.4	26.8	589	11	BG793616	BG793616
C 15	32	26.4	266	10	BB608462	BB608462
C 16	32	26.4	474	13	AZ049530	AZ049530
C 17	32	26.4	555	10	AF657239	AF657239
C 18	32	26.4	842	11	BF696719	BF696719
C 19	32	26.4	952	13	AQ744189	AQ744189
C 20	32	26.4	1069	11	BF871820	BF871820
C 21	31.8	26.3	728	11	BF578750	BF578750
C 22	31.4	26.0	218	10	AA055704	AA055704
C 23	31.4	26.0	285	11	F11815	F11815
C 24	31.4	26.0	436	10	AA511221	AA511221
C 25	31.4	26.0	692	10	BE785913	BE785913
C 26	31.4	26.0	1002	11	BG328641	BG328641
C 27	31.2	25.8	262	13	B47161	B47161
C 28	31.2	25.8	458	11	BF063108	BF063108
C 29	31.2	25.8	473	13	AQ332734	AQ332734
C 30	31.2	25.8	607	13	AZ593838	AZ593838
C 31	31.2	25.8	648	11	B1227246	B1227246
C 32	31.2	25.8	961	11	BF621610	BF621610
C 33	31.2	25.8	1293	11	BF864875	BF864875
C 34	31	25.6	669	13	AQ483120	AQ483120
C 35	31	25.6	1068	10	BE512646	BE512646
C 36	31	25.6	1110	10	BE787988	BE787988
C 37	31	25.6	1187	11	BF342437	BF342437
C 38	31	25.6	1201	13	CNS010AP	CNS010AP
C 39	30.8	25.5	244	10	BE163170	BE163170
C 40	30.8	25.5	274	10	BB565641	BB565641
C 41	30.8	25.5	309	10	AA623926	AA623926
C 42	30.8	25.5	354	11	W97467	W97467
C 43	30.8	25.5	374	10	AA036350	AA036350
C 44	30.8	25.5	377	10	AA049789	AA049789
C 45	30.8	25.5	403	11	W12086	W12086

ALIGNMENTS

RESULT 1	AA311787	273 bp	mRNA	EST	19-APR-1997
LOCUS	EST182503	Jurkat T-cells	VI	Homo sapiens	cDNA 5' end similar to
DEFINITION	similar to T-cell factor 1, A/B/C, mRNA sequence.				
ACCESSION	AA311787				
VERSION	AA311787.1	GI:1964114			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	1 (bases 1 to 273)				
	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult				
	,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White				
	,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,				
	Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald				
	,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A.,				
	Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,				
	Kelley,J.C., Liu,L.-L., Marmaros,S.M., Merrick,J.M.,				
	Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,				
	Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,				
	Small,K.V., Spriggs,T.A., Ulterback,T.R., Weidman,J.F., Li,Y.,				
	Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,				
	Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.				
	,Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,				
	Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,				
	Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon				
	,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and				
	Venter,J.C.				
TITLE	Initial assessment of human gene diversity and expression patterns				
JOURNAL	based upon 83 million nucleotides of cDNA sequence				
MEDLINE	Nature 377 (6547 Suppl), 3-174 (1995)				
COMMENT	96026280				
	Contact: Kerlavage, AR				

Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse

FEATURES

source

1. .273
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="ATCC (inhost):158998"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells VI"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 60 a 111 c 59 g 39 t 4 others
ORIGIN

Query Match 85.1%; Score 103; DB 10; Length 273;
Best Local Similarity 98.1%; Pred. No. 1.8e-15;
Matches 103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 accagcgcatgacaaagagaccgtctactccgccttcctcaatgtctcatgcatcaccaca 60
|||||
Db 77 ACCAGCGCATGTACAAAGAGACCGTCTACTCCGCTTCAATGCTCATGTCATTACCCN 136
Oy 61 cccctctcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 105
|||||
Db 137 NCCCTCGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 181

RESULT 2

LOCUS BG386035 353 bp mRNA EST 12-MAR-2001
DEFINITION 602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5',
mRNA sequence.
ACCESSION BG386035.1 GI:13279481
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 353)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LUCM1307 row: h column: 02
High quality sequence stop: 276.
Location/Qualifiers

FEATURES

source

1. .353
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 80 a 111 c 122 g 40 t
ORIGIN

Query Match 73.1%; Score 88.4; DB 11; Length 353;
Best Local Similarity 98.9%; Pred. No. 5.9e-12;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 accagcgcatgtacaaagagaccgtctactccgccttcctcaatgtctcatgcatcaccaca 60
|||||
Db 193 ACCAGCGCATGTACAAAGAGACCGTCTACTCCGCTTCAATGCTCATGTCATTACCCA 252
Oy 61 cccctctcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 90
|||||
Db 253 CCCCCCTCGGGAGCAGGGCAGCAGCAGCAGCAGCAGCAG 282

RESULT 3

LOCUS BG655601 400 bp mRNA EST 05-JUL-2001
DEFINITION la84402.x1 Melton Mouse E16.5 Pancreas Library M1621 Mus musculus
cDNA 3' similar to SW:TCF1_MOUSE Q00417 T-CELL-SPECIFIC
TRANSCRIPTION FACTOR 1; mRNA sequence.
ACCESSION BG655601
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
Schmitt, A., Theising, B., Ritter, K., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -400P from Gibco
High quality sequence stop: 394.
Location/Qualifiers

FEATURES

source

1. .400
Location/Qualifiers
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone_lib="Melton Mouse E16 5 Pancreas Library M1621"
/sex="Both"
/tissue_type="Total pancreas"
/dev_stage="Embryonic day 16.5"
/lab_host="TOP10"
/note="Organ: Pancreas; Vector: pZErO-2; Site_1: NotI;
Site_2: XhoI; Library constructed using SuperScript
Plasmid Library kit (Life Technologies). cDNA made by

BASE COUNT	45 a	71 c	50 g	35 t
ORIGIN				
Query Match	54.5%	Score 66;	DB 11;	Length 201;
Best Local Similarity	100.0%;	Pred. No. 1.4e-06;		
Matches 66;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT	7
AI049127/c	
LOCUS	AI049127 288 bp EST 08-JUL-1998
DEFINITION	ub35d04.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:1379719 5' similar to gb:X59869 TRANSCRIPTION FACTOR-7 (HUMAN)); gb:X61385 Mouse mRNA for T-cell specific transcription factor (MOUSE):: mRNA sequence.

ACCESSION	AI049127	
VERSION	AI049127.1	GI:3297414
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
REFERENCE	1 (bases 1 to 288)	

AUTHORS	TITLE	JOURNAL	COMMENT
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouquet, L., Gaisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morriss, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, R., Wilson, R. and Waterston, R.	The WashU-BHI Mouse EST Project Unpublished (1996) Contact: Marra M/Mouse EST project	WashU-BHI Mouse EST project Washington University School of Medicine	

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNI; contact the
IMAGE Consortium (infoimage.llni.gov) for further information.
MGI:902187
Seq primer: -28m13 rev2 ET from Amersham.
Location/Qualifiers
1..288
FEATURES
source

```

/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1379719"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="vector: p77T3D-pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand
was primed with a Not I - oligo(dT) primer [5,

```

BASE COUNT	ORIGIN
56 a	84 c
117 g	73 t

Query Match	51.9%	Score 62.8;	DB 10;	Length 330;
Best Local Similarity	75.8%;	Pred. No. 8.7e-06;		
Matches 91; Conservative	0;	Mismatches 27;	Indels 2;	Gaps 1;

Db CCAGTGGCATGTACAAAGACTGCTACTCGCTCAATCTGCTCATGCCCTACCCAC 127
Qy ccccttcggagcaggcgagcaccccacgcgcgccccgcgttgtaagtgaaccgcc 121
Db CGGCTCCGS - A&A:AGGCAGCATCGCAGGCTCAACCCGCTGCATAACAAGCAGCGCC 69

RESULT	9	EST	14-SEP-1998
LOCUS	AI131784	152 bp	mRNA
DEFINITION	AI131784	uc37a06.rl Soares_mammary_gland_NbMMG	Mus musculus cDNA clone IMAGE:1400146 5', similar to gb:X61385 Mouse mRNA for T-cell specific transcription factor (MOUSE); mRNA sequence.

EST.
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Fukuyota; Metazoa; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 152)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Thelsing, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

Unpublished (1990).
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLMU : contact the
IMAGE Consortium (info@image.llnu.gov) for further information.
MGI:911862
Trace considered overall poor quality
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence set:1.

```
i..152
source
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/image="IMAGE:I400146"
/clone_lib="Scares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note-organ:"mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco R; 1st strand cDNA was primed with a Nol I - oligo(dT) primer 15' TGCTTACCAATCTGAAGTCGGAGCGGCAGCAATTGTTTTTTTTTTTT
```

Db	258	CCAGGGAGTGGAGAAATTATCTGTGTCATATTGGAGTTCAGAGTCTCTCGATTACGAA	317
Qy	62	ccccctcgggagcggcagcagccccagccgcagccccgcctggtaagtyggaccccy	119
Db	318	AACCTCTGCAGCTTGCTGCTCCCAACAGCTGCCAGTTTGTCTCTTGAACCCG	375

LOCUS	BF538482	587 bp	mRNA	EST	11-DEC-2000
DEFINITION	602053985F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:419301 mRNA sequence.				

ACCESSION
BF538482
VERSION
BF538482.1
GI:11625850

KEYWORDS
SOURCE

ORGANISM

REFERENCE

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey F. Green, M.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the MGC Consortium (<http://www.mgc.org>)

http://image.llnl.gov
plate: LLAM9524 row: j column: 21
High quality sequence stop: 587.

```

1. .567
source
/organism="Mus musculus"
/strain="FVB/N"
db xref="taxon.10090"

```

```

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: C

```

Query Match 27.9%; Score 33.8; DB 11; Length 587;

Qy 26 tctactcgcgccttcaatctgctcatgcatcaccacccccctcgg 70
||||| ||||||| ||||||| ||||||| |||
Db 1 TCTACTCTGGCTCAATCTGCTCATGCGCTACCCACCGGTTCCGG 45

LOCUS	BF846035	411 bp	mrna	EST	16-JAN-2001
DEFINITION	PM1-EN0060-151000-001-d11 EN0060 Homo sapiens cDNA, mRNA sequence				
ACCESSION	BF846035				
VERSION	BF846035.1				

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

REFERENCE
AUTHORS
1 (bases 1 to 411)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.,
Neves, M. A. da Silva, U. et al. (2005) The genome of the

Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermoactivation of thermostable enzymes by
trichalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

BASE COUNT	42 a	37 c	51 g	61 t	1 others
ORIGIN					
Query Match		26.9%	Score 32.6;	DB 10;	Length 192;
Best Local Similarity		55.4%	Pred. No. 1.6e+02;		
				Models 50;	Gaps 0;

ORIGIN	Query Match	Score 32.6;	DB 10;	Length 192;
	Best Local Similarity	55.4%;	Pred. No. 1.6e+02;	
				0: Gaps
				Models

Db 52 CCGCCCTACCAACAGGAATCTCCCCCGGAGCCCGCGCTTGGTGTTG 1

RESULT 14
BG793616/c

16 MAY 2001

KEYWORDS
 house mouse.
SOURCE
 Mus musculus
ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE
 1 (bases 1 to 589)
AUTHORS
 Gallardo, T.D., Schaghaman, J.J., Pertsemilidis, A., Garner, H.R.,
 Williams, R.S., and Schobert, R.V.

TITLE	UT Southwestern Medical Center, Adult Mouse Skeletal Muscle
JOURNAL	Unpublished (2001)
COMMENT	Contact: Schageman JJ Shohet/Garner Labs UT Southwestern Medical Center

FEATURES
 Source
 1. 589
 location/Qualifiers
 Seq primer: M13/pUC reverse.
 sequencing; ABI Prism 377 sequencer and analysis software.
 us base calls or 'N's in windowed segments. Sequencing: First-pass
 Quality: Sequence ends were trimmed based on percentage of ambigu
 Prevention grants for use in cDNA microarray experiments. Sequence
 Genomic Applications (PGA) and the Reynolds Heart Disease
 cDNA library constructed by UTSW as a component of the program for
 Email: Jeff_Schageman@UTSouthwestern.edu

us-09-904-420a-1_copy_820_940.rst

Thu Feb 21 07:25:38 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 11:02:39 ; Search time 1729.17 Seconds
(without alignments)
171.729 Million cell updates/sec

Title: US-09-904-420A-9

Perfect score: 18

Sequence: 1 ccgaagtgggtgggtaaat 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.sy.*

28: em.un.*

29: em.vi.*

30: em.htgo_hum.*

31: em.htgo_inv.*

32: em.htgo_rod.*

33: em.htg_hum.*

34: em.htg_inv.*

35: em.htg_rod.*

36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18	100.0	202915	2	AC060773	AC060773 Homo sapi
C 2	17	94.4	72449	9	AC005658	AC005658 c1b_11_0
C 3	17	94.4	73522	2	AC006173	AC006173 Homo sapi
C 4	17	94.4	116058	9	AC005659	AC005659 c1b_161_
C 5	17	94.4	122014	9	AL359833	AL359833 Human DNA
C 6	17	94.4	126232	9	AL356419	AL356419 Human DNA
C 7	17	94.4	172038	2	AC025425	AC025425 Homo sapi
C 8	17	94.4	199008	2	AC091699	AC091699 Homo sapi
C 9	16.4	91.1	251	11	G66013	G66013 sy890 Misce
C 10	16.4	91.1	1115	9	HSTCF1D	Z47361 H.sapiens T
C 11	16.4	91.1	1165	9	HSTCF1E	Z47362 H.sapiens T
C 12	16.4	91.1	1254	9	HSTCF1A	X59869 Human TCF-1
C 13	16.4	91.1	2814	9	HSTCF1B	X59870 Human TCF-1
C 14	16.4	91.1	2855	9	HSTCF1G	X63901 Homo sapien
C 15	16.4	91.1	2910	9	HSTCF1C	X59871 Human TCF-1
C 16	16.4	91.1	45896	9	HSBA9F11	AL050312 Human DNA
C 17	16.4	91.1	61041	2	AC074135	AL133231 Human DNA
C 18	16.4	91.1	65756	2	AC068850	AC074135 Homo sapi
C 19	16.4	91.1	67331	2	AC010286	AC068850 Homo sapi
C 20	16.4	91.1	71842	2	AC036189	AC010286 Homo sapi
C 21	16.4	91.1	76094	9	HS179D3A	Z81364 Human DNA s
C 22	16.4	91.1	80154	2	AC011502	AC011502 Homo sapi
C 23	16.4	91.1	80568	9	AC002420	AC002420 Human Chr
C 24	16.4	91.1	83030	9	AC004814	AC004814 Homo sapi
C 25	16.4	91.1	83120	9	HS796117	AL035398 Human DNA
C 26	16.4	91.1	84544	2	AC009012	AC009012 Homo sapi
C 27	16.4	91.1	86914	2	AC011336	AC011336 Homo sapi
C 28	16.4	91.1	95445	2	AC016175	AC016175 Homo sapi
C 29	16.4	91.1	100316	9	AC004556	AC004556 Homo sapi
C 30	16.4	91.1	102672	9	AC013652	AC013652 Homo sapi
C 31	16.4	91.1	106827	9	AC004929	AC004929 Homo sapi
C 32	16.4	91.1	108210	2	AL158845	AL158845 Homo sapi
C 33	16.4	91.1	115431	9	AC010880	AC010880 Homo sapi
C 34	16.4	91.1	115855	9	AC012611	AC012611 Homo sapi
C 35	16.4	91.1	115932	9	AC011446	AC011446 Homo sapi
C 36	16.4	91.1	117229	9	AC021083	AC021083 Homo sapi
C 37	16.4	91.1	119185	9	AL157699	AL157699 Human DNA
C 38	16.4	91.1	120829	2	AC009829	AC009829 Homo sapi
C 39	16.4	91.1	123160	2	AC083773	AC083773 Homo sapi
C 40	16.4	91.1	123551	2	AC005809	AC005809 Homo sapi
C 41	16.4	91.1	124347	9	AC010072	AC010072 Homo sapi
C 42	16.4	91.1	129414	2	AC068398	AC068398 Homo sapi
C 43	16.4	91.1	130754	2	AC008508	AC008508 Homo sapi
C 44	16.4	91.1	130767	9	AL161724	AL161724 Human DNA
C 45	16.4	91.1	130767	9	AL161724	AL161724 Human DNA

ALIGNMENTS

RESULT 1
AC060773/c

LOCUS

DEFINITION

SEQUENCE, 8 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC060773 202915 bp DNA HTG 09-MAY-2001
Homo sapiens chromosome 4 clone RP11-741G21 map 4, WORKING DRAFT
SEQUENCE, 8 unordered pieces.

AC060773 GI:12061518

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 202915)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 4, clone RP11-741G21

Unpublished

2 (bases 1 to 202915)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Beldwin,J., Barna,N., Rastien,V., Beda,F.,

Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kaun,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (20-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 10, 2001 this sequence version replaced gi:7798794.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7430
 Center clone name: 741_G_21

----- Summary Statistics
 Sequencing vector: M13; M77815; 49% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 19511 bases at least Q40
 Consensus quality: 201020 bases at least Q30
 Consensus quality: 201685 bases at least Q20
 Insert size: 194000; agarose-fp
 Insert size: 202215; sum-of-contigs
 Quality coverage: 9.5 in Q20 bases; agarose-fp
 Quality coverage: 9.1 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 3959; contig of 3959 bp in length
 * 3960 4059; gap of 100 bp
 * 4060 5548; contig of 1489 bp in length
 * 5549 5648; gap of 100 bp
 * 5649 7747; contig of 2099 bp in length
 * 7748 7847; gap of 100 bp
 * 7848 75665; contig of 67818 bp in length
 * 75666 75765; gap of 100 bp
 * 75766 109526; contig of 33761 bp in length
 * 109527 109626; gap of 100 bp
 * 109627 14431; contig of 34805 bp in length
 * 14432 144531; gap of 100 bp
 * 144532 184185; contig of 39654 bp in length
 * 184186 184285; gap of 100 bp
 * 184286 202915; contig of 18630 bp in length.

Location/Qualifiers
 1..202915
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"

FEATURES
 source

/map="4"
 /clone="rp11-741G21"
 /clone_lib="RPC1-II Human Male BAC"
 1..3959
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 4060..5548
 /note="assembly_fragment"
 5649..7747
 /note="assembly_fragment"
 7848..75665
 /note="assembly_fragment"
 75766..109526
 /note="assembly_fragment"
 109627..14431
 /note="assembly_fragment"
 144532..184185
 /note="assembly_fragment"
 184286..202915
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"
 BASE COUNT 61574 a 37018 c 37733 g 65883 t 707 others
 ORIGIN
 Query Match 100.0%; Score 18; DB 2; Length 202915;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccgaggtgggtgggtaat 18
 Db 91305 CCGAGGTGGGTGGGTAAT 91288
 RESULT 2
 AC005658/ 72449 bp DNA PRI 05-NOV-1999
 LOCUS
 c1tb.ll.o_6, complete sequence.
 AC005658
 AC005658.3 GI:6249679
 VERSION
 HTG.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 72449)
 Smith,D.R.
 TITLE
 Sequencing of Human Chromosome 10
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 72449)
 Smith,D.R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (10-SEP-1998) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02154, USA
 REFERENCE
 3 (bases 1 to 72449)
 Smith,D.R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (31-OCT-1998) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02154, USA
 REFERENCE
 4 (bases 1 to 72449)
 Smith,D.R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02154, USA
 REFERENCE
 5 (bases 1 to 72449)
 Smith,D.R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02154, USA
 REMARK
 Vector Sequence clipped
 COMMENT
 On Nov 5, 1999 this sequence version replaced gi:4314336.

FEATURES	source		Location/Qualifiers	
			1..72449 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /map="10q25" /clone_lib="CIT987SK-101106"	
BASE COUNT	18794 a	15799 c	16718 g	21138 t
ORIGIN				
Query Match	94.4%; Score 17; DB 9; Length 72449;			
Best Local Similarity	100.0%; Pred. No. 90;			
Matches	17; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	1	ccgaggtgggtgggttaa	17	
Db	35365	CCGAGTGGGTGGGTAA	35349	
RESULT 3				
AC006173/c				
LOCUS				
DEFINITION	AC006173 73522 bp DNA HTG 09-DEC-1998			
ACCESSION	Homo sapiens chromosome 10 clone CIT987SK-1161C12 map 10q25, ***			
VERSION	SEQUENCING IN PROGRESS ***, 3 ordered pieces.			
KEYWORDS	AC006173			
SOURCE	HTG: HTGS_PHASE2.			
ORGANISM	human.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	Smith,D.R.			
JOURNAL	Sequencing of Human Chromosome 10			
REFERENCE	Smith,D.R.			
AUTHORS	Unpublished			
TITLE	2 (bases 1 to 73522)			
JOURNAL	Smith,D.R.			
COMMENT	Submitted (09-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
FEATURES	Technology. Data may contain low quality seq uence and BAC/Cosmid vector sequences.			
source	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 3 contigs. Gaps between the contigs			
	* are represented as runs of N. The order of the pieces			
	* is believed to be correct as given, however the sizes			
	* of the gaps between them are based on estimates that have			
	* provided by the submittor.			
	* This sequence will be replaced			
	* by the finished sequence as soon as it is available and			
	* the accession number will be preserved.			
	* 1 2074: contig of 2074 bp in length			
	* 2075 4552: contig of 2478 bp in length			
	* 4553 73522: contig of 68970 bp in length.			
FEATURES	Location/Qualifiers			
source	1..73522 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CIT987SK-1161G12" /chromosome="10" /map="10q25"			
BASE COUNT	19125 a	16131 c	16684 g	21561 t
ORIGIN	19125 a 16131 c 16684 g 21561 t 21 others			
Query Match	94.4%; Score 17; DB 2; Length 73522;			
Best Local Similarity	100.0%; Pred. No. 90;			
Matches	17; Conservative	0; Mismatches	0; Indels	0; Gaps

Qy	1	ccgaggtgggtgggttaa	17	
Db	30353	CCGAGTGGGTGGGTAA	30337	
RESULT 4				
AC005659				
LOCUS				
DEFINITION	AC005659 116058 bp DNA PRI 05-NOV-1999			
ACCESSION	C11b.161.g_12, complete sequence.			
VERSION	AC005659			
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 116058)			
JOURNAL	Smith,D.R.			
REFERENCE	Sequencing of Human Chromosome 10			
AUTHORS	Unpublished			
TITLE	2 (bases 1 to 116058)			
JOURNAL	Smith,D.R.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (10-SEP-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
TITLE	3 (bases 1 to 116058)			
JOURNAL	Smith,D.R.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
TITLE	4 (bases 1 to 116058)			
JOURNAL	Smith,D.R.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
TITLE	5 (bases 1 to 116058)			
JOURNAL	Smith,D.R.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
TITLE	Vector Sequence Clipped			
JOURNAL	On Nov 5, 1999 this sequence version replaced gi:4314333.			
COMMENT	Location/Qualifiers			
FEATURES	1..116058 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /map="10q25" /clone_lib="CIT987SK-1161G12" /clone="10q25"			
BASE COUNT	34107 a	26479 c	25708 g	29764 t
ORIGIN				
Query Match	94.4%; Score 17; DB 9; Length 116058;			
Best Local Similarity	100.0%; Pred. No. 82;			
Matches	17; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	1	ccgaggtgggtgggttaa	17	
Db	88217	CCGAGTGGGTGGGTAA	88233	
RESULT 5				
AL359833				
LOCUS				
DEFINITION	AL359833 122014 bp DNA PRI 23-NOV-2000			
ACCESSION	Human DNA sequence from clone RP11-186C9 on chromosome 1, complete sequence.			
VERSION	AL359833			
KEYWORDS	AL359833.12 GI:11340294			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

repeat_region	31188..31219	/note="8 copies 4 mer gtgt 87% conserved"
repeat_region	32209..32257	/note="L2 repeat: matches 2702..2750 of consensus"
repeat_region	32281..32339	/note="MIR repeat: matches 110..168 of consensus"
repeat_region	32367..32477	/note="MIR repeat: matches 148..262 of consensus"
repeat_region	34251..34437	/note="MER5A repeat: matches 1..189 of consensus"
misc_feature	34326..34719	/note="match: GSS: Em:AQ186533"
repeat_region	35098..35173	/note="MIR repeat: matches 59..145 of consensus"
repeat_region	35418..35735	/note="AluSq repeat: matches 4..311 of consensus"
repeat_region	36438..36483	/note="23 copies 2 mer ga 80% conserved"
repeat_region	36440..36483	/note="L1 copies 4 mer gaga 81% conserved"
repeat_region	36691..37152	/note="MLTIC repeat: matches 1..466 of consensus"
repeat_region	37163..37465	/note="AluDb repeat: matches 1..295 of consensus"
repeat_region	37702..37869	/note="MLTIF repeat: matches 56..233 of consensus"
repeat_region	37926..38145	/note="MLTIF repeat: matches 232..460 of consensus"
misc_feature	38432..38480	/note="match: GSS: Em:AQ123310"
repeat_region	38734..39111	/note="MLTIF repeat: matches 485..533 of consensus"
misc_feature	39817..39968	/note="match: GSS: Em:AQ278876"
repeat_region	40082..40232	/note="L2 repeat: matches 2591..2750 of consensus"
repeat_region	41093..41206	/note="L2 repeat: matches 2310..2470 of consensus"
repeat_region	41871..42064	/note="38 copies 3 mer gga 90% conserved"
repeat_region	42275..42694	/note="L1M4 repeat: matches 5148..5344 of consensus"
repeat_region	43257..43288	/note="L1M4 repeat: matches 3812..4238 of consensus"
Query Match 94.4%; Score 17; DB 9; Length 122014;		
Best Local Similarity 100.0%; Pred. No. 81;		
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	cagagtggtgggttaa 17
Db	37210	CCGAGGTGGGTGGTAA 37226
RESULT	6	
LOCUS	AL356419/c	126232 bp DNA PRI 18-JUL-2001
DEFINITION	Human DNA sequence from clone RP11-99L6 on chromosome 10, complete sequence.	
ACCESSION	AL356419	
VERSION	AL356419.10	GI:14970798
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 126232)	
JOURNAL	Sanger Centre, Hinxton, Cambridgeshire, UK. E-mail enquiries: humquery@sanger.ac.uk	

COMMENT	On Jul 19, 2001 this sequence version replaced gi:14268075. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one W13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10 RP11-99L6 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACet.6 IMPORTANT: This sequence is not the entire insert of clone RP11-99L6 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-99L6 is at 126232 in this sequence. The true left end of clone RP11-328K15 is at 107108 in this sequence. The true right end of clone RP11-140G2 is at 100 in this sequence.	
FEATURES	Location/Qualifiers	
source	1..126232	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="10"
		/clone="RP11-99L6"
		/clone_lib="RPCI-11.1"
repeat_region	832..1138	/note="AluX repeat: matches 1..307 of consensus"
repeat_region	1184..1306	/note="FIAM_C repeat: matches 4..124 of consensus"
repeat_region	1495..1600	/note="L2 repeat: matches 2579..2690 of consensus"
repeat_region	1608..1901	/note="AluJo repeat: matches 17..307 of consensus"
repeat_region	2656..2726	/note="MLT1 repeat: matches 72..143 of consensus"
repeat_region	2864..2959	/note="MLT1H repeat: matches 406..493 of consensus"
repeat_region	2960..3402	/note="L1PA10 repeat: matches 5710..6161 of consensus"
repeat_region	3403..3442	/note="MLT1H repeat: matches 493..534 of consensus"
repeat_region	3737..3932	/note="MLT1H repeat: matches 116..318 of consensus"
repeat_region	4006..4092	/note="MLT1H repeat: matches 456..547 of consensus"
repeat_region	4129..4195	/note="L2 repeat: matches 2685..2748 of consensus"
repeat_region	4547..4688	/note="MIR repeat: matches 56..207 of consensus"
repeat_region	5244..5526	/note="L2 repeat: matches 1897..2160 of consensus"
repeat_region	5527..5837	/note="AluSq repeat: matches 1..311 of consensus"
repeat_region	5838..6355	/note="L2 repeat: matches 2160..2700 of consensus"
repeat_region	8478..9427	/note="MLT2B repeat: matches 1..448 of consensus"

repeat_region	9428. .9461 /note="MIR repeat: matches 158. .192 of consensus"
repeat_region	10349. .10379 /note="LIM3 repeat: matches 5486. .5516 of consensus"
repeat_region	10814. .10742 /note="MIR repeat: matches 121. .262 of consensus"
repeat_region	10806. .11101 /note="AluJo repeat: matches 18. .304 of consensus"
repeat_region	11173. .12118 /note="LIMC4 repeat: matches 7007. .7977 of consensus"
repeat_region	12193. .12399 /note="MIR repeat: matches 17. .216 of consensus"
repeat_region	12404. .12636 /note="MIR repeat: matches 9. .257 of consensus"
repeat_region	13697. .13917 /note="LIR33 repeat: matches 291. .514 of consensus"
repeat_region	14018. .14117 /note="LIR33 repeat: matches 33. .142 of consensus"
repeat_region	14249. .14332 /note="MER91A repeat: matches 74. .158 of consensus"
repeat_region	14722. .14927 /note="MIR repeat: matches 38. .262 of consensus"
repeat_region	15428. .15744 /note="L2 repeat: matches 2402. .2709 of consensus"
repeat_region	15956. .16261 /note="AluJb repeat: matches 1. .306 of consensus"
repeat_region	19777. .20064 /note="AluJo repeat: matches 1. .282 of consensus"
repeat_region	20086. .20237 /note="LIME3 repeat: matches 5706. .5850 of consensus"
repeat_region	22131. .22314 /note="MIR repeat: matches 61. .250 of consensus"
misc_feature	24244 /note="Weak data."
repeat_region	24414. .24722 /note="AluSc repeat: matches 1. .305 of consensus"
repeat_region	26368. .26652 /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region	26929. .27019 /note="MIR repeat: matches 80. .177 of consensus"
repeat_region	27069. .27349 /note="AluSx repeat: matches 29. .312 of consensus"
repeat_region	27461. .27514 /note="L2 repeat: matches 2558. .2710 of consensus"
repeat_region	28932. .29007 /note="MIR repeat: matches 30. .100 of consensus"
repeat_region	29311. .29357 /note="Alu repeat: matches 255. .301 of consensus"
repeat_region	29427. .29639 /note="MIR repeat: matches 28. .251 of consensus"
repeat_region	29595. .29655 /note="L2 repeat: matches 2649. .2709 of consensus"
repeat_region	29952. .30450 /note="LIMB3 repeat: matches 5678. .6184 of consensus"
repeat_region	30468. .30513 /note="23 copies 2 mer at 76% conserved"
repeat_region	31081. .31210 /note="MIR repeat: matches 5. .147 of consensus"
repeat_region	32141. .32152 /note="L2 repeat: matches 2732. .2744 of consensus"
repeat_region	32153. .32439 /note="AluSx repeat: matches 12. .297 of consensus"
repeat_region	32440. .32971 /note="L2 repeat: matches 2256. .2732 of consensus"
repeat_region	33063. .33253 /note="MER58A repeat: matches 2. .80 of consensus"
repeat_region	33254. .33479 /note="AluJy repeat: matches 76. .299 of consensus"
repeat_region	33480. .33506 /note="MER58A repeat: matches 80. .214 of consensus"
repeat_region	33507. .33646 /note="MIR repeat: matches 7. .151 of consensus"
repeat_region	33772. .33845

repeat_region	/note="MIR repeat: matches 80. .154 of consensus" 34884. .35077 /note="L1PB1 repeat: matches 5951. .6155 of consensus" 35780. .35827 /note="MIR repeat: matches 103. .151 of consensus" 36408. .36552 /note="MLT1C repeat: matches 6. .144 of consensus" 36553. .37105 /note="MER41A repeat: matches 1. .554 of consensus" 37106. .37394 /note="MLT1C repeat: matches 144. .458 of consensus" 38198. .38233 /note="6 copies 6 mer tgtgtg 97% conserved" 38693. .38883 /note="MIR repeat: matches 16. .251 of consensus" 38921. .39105 /note="MIR repeat: matches 13. .200 of consensus" 39331. .39400 /note="35 copies 2 mer ac 67% conserved" 40666. .40751 /note="MER5B repeat: matches 1. .83 of consensus" 40772. .40812 /note="MER5B repeat: matches 1. .41 of consensus" 41207. .41379 /note="MIR repeat: matches 27. .214 of consensus" 41836. .42384 /note="L1ME3 repeat: matches 5618. .6162 of consensus" 42412. .42688 /note="L1ME3 repeat: matches 5261. .5548 of consensus" 42842. .43224 /note="L1ME3 repeat: matches 4801. .5197 of consensus" 43433. .43799 /note="THE1C repeat: matches 1. .370 of consensus" 43802. .44018 /note="MIR repeat: matches 3. .248 of consensus" 44062. .44190 /note="MIR repeat: matches 25. .151 of consensus" 45135. .45436	94.4%; Score 17; DB 9; Length 126232; Best Local Similarity 100.0%; Pred. No. 81; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ccgaggtgggtggttaa 17 	
Pb	24576 CCAGGTGGTGGGTAA 24660 	

Query Match 94.4%; Score 17; DB 9; Length 126232;
Best Local Similarity 100.0%; Pred. No. 81;

Qy 1 ccgagtggtggtaa 17
|||||
pb 24676 CCGAGTGGGTGGTAA 24660

```

RESULT 7
AC025425 172038 bp DNA HTG 17-JUL-2001
LOCUS Homo sapiens chromosome 10 clone RP11-178C16, WORKING DRAFT
DEFINITION SEQUENCE, 7 unordered pieces.
AC025425
AC025425.9 GI:14787169
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172038)
AUTHORS Smith, D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
JOURNAL 2 (bases 1 to 172038)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Jul 17, 2001 this sequence version replaced gi:14277218.
----- Genome Center
Genome Therapeutics Corporation

```



```
misc_feature 2661..6272
              /note="assembly_name:Contig17
              clone_end:T7"
              6373..11483
              /note="assembly_name:Contig18"
              11584..17593
              /note="assembly_name:Contig19"
              17694..25191
              /note="assembly_name:Contig20"
              25292..35393
              /note="assembly_name:Contig21"
              35494..62376
              /note="assembly_name:Contig22"
              62477..103800
              /note="assembly_name:Contig23
              clone_end:SP6"
              103901..199008
              /note="assembly_name:Contig24"
              63479 a 38138 c 36348 g 60242 t 801 others
              ORIGIN

Query Match          94.4%; Score 17; DB 2; Length 199008;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cccagggtgggtgggtaa 17
    |||
Db 2131 CCGAGGTGGGTGGGTAA 2147

RESULT 9
HSTCFID/c
LOCUS          G66013          251 bp      DNA          STS          01-AUG-2001
DEFINITION     sy890 Miscellaneous Y sequences Homo sapiens STS genomic, sequence
               tagged site.
ACCESSION      G66013
VERSION        G66013.1 GI:15078064
KEYWORDS       STS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 251)
AUTHORS        Tilford,C.A., Kuroda-Kawaguchi,T., Skaletsky,H., Rozen,S.,
               Brown,L.G., Rosenber,M., McPherson,J.D., Wylie,K., Sekhon,M.,
               Kucaba,T.A., Waterston,R.H. and Page,D.C.
TITLE          The human Y chromosome: A mosaic of gene-laden amplicons and
               X-homologous domains
JOURNAL        Unpublished
COMMENT        Synonyms: PRKY
               Contact: Tomoko Kawaguchi
               Page Lab
               Whitehead Institute for Biomedical Research
               Room 423, 9 Cambridge Center, Cambridge, MA 02142, USA
               Email: kawaguchi@wi.mit.edu
               Primer A: CCGAGTGGGTGGTAATTTG
               Primer B: TTTTCTGGTGGAGTCGCT
               STS size: 244
               PCR Profile:
               94C          3:00 min
                   / 94C          1:00 sec
               35x 1 60C          1:00 min
                   \ 72C          1:00 min
               72C          5:00 min

Protocol:
ng DNA
Template: 1 ul saturated bacterial culture (BACs) or 100
Primer:  each 1 uM
dNTPs:  each 100 uM
Tag Polymerase: 0.05 units/ul
Total Vol: 15 ul or 20 ul
Mineral Oil: 15 ul

Buffer:
For 20ml of 10x Stock solution:
Stock Reagent      Volume Mixed
1M KCl              10 ml
1M Tris-HCl, pH 9.0 2 ml
Triton X-100        200 ul
2M MgCl2            150 ul
H2O                 7.65 ml.

Final Conc. of 10x Stock
500 mM
100 mM
1.0 %
15 mM

FEATURES
source
1..251
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Miscellaneous Y sequences"
/note="Human STSs derived from miscellaneous Y sequences"
STS
2..245
primer_bind 2..21
primer_bind 83 a complement(226..245) 45 t
BASE COUNT 83 a 58 c 65 g
ORIGIN

Query Match          91.1%; Score 16.4; DB 11; Length 251;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 cccagggtgggtgggtaat 18
    |||
Db 1 CCGAGGTGGGTGGGTAAT 18

RESULT 10
HSTCFID/c
LOCUS          HSTCFID          1115 bp      mRNA          PRI          09-JAN-1995
DEFINITION     H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.
ACCESSION      247361
VERSION        247361.1 GI:619881
KEYWORDS       splice form D; T cell factor 1.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1115)
AUTHORS        Mayer,K., Wolff,E., Clevers,H. and Ballhausen,W.G.
TITLE          The human high mobility group (HMG)-box transcription factor TCF-1:
               novel isoforms due to alternative splicing and usage of a new exon
JOURNAL        IXA
COMMENT        Unpublished
               2 (bases 1 to 1115)
               van de Wetering,M., Oosterwegel,M., Holstege,F., Dooyes,D.,
               Suijkerbuijk,R., Geurts van Kessel,A. and Clevers,H.
               The human T cell transcription factor-1 gene. Structure,
               localization, and promoter characterization
               J. Biol. Chem. 267 (12), 8530-8536 (1992)
               92235082
               3 (bases 1 to 1115)
               Ballhausen,W.G.
               Direct Submission
               Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
               Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
               Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES
source
1..1115
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1..1115
/feature="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/function="high mobility group box transcription factor"
```

/evidence=experimental		/function="high mobility group box transcription factor"			
gene	1. .1115	/evidence=experimental			
CDS	/gene="TCF-1"	1. .1165			
	2. .754	2. .1144			
/gene="TCF-1"		/standard_name="T cell factor 1 splice form D"			
/citation=[1]		/standard_name="T cell factor 1 splice form E"			
/codon_start=1		/citation=[1]			
/product="T cell factor 1 splice form D"		/codon_start=1			
/protein_id="CAA87439.1"		/product="T cell factor 1 splice form E"			
/db_xref="GI:619882"		/protein_id="CAA87440.1"			
/translation="MYKETVYSANFLMLHYPPSPGAGHPQPPLHKANOPPHGVQ		/db_xref="GI:619884"			
LSLVEHNSPHPTAPADISOKOVHRLPLOTDLGFSYLTSGSMGQLPHTVSWFTHPS		/translation="MYKETVYSANFLMLHYPPSPGAGHPQPPLHKANOPPHGVQ			
LMLGSGVGHGPAALPHPAIVPPSGKOELQDFDRLNLTQAEKAEKAKPTIKKPLNA		LSLVEHNSPHPTAPADISOKOVHRLPLOTDLGFSYLTSGSMGQLPHTVSWFTHPS			
FMLTWKEMRAKYTAECTILKESAINIOLGRWHALSREOAKYVELARKERQLHMQLY		LMLGSGVGHGPAALPHPAIVPPSGKOELQDFDRLNLTQAEKAEKAKPTIKKPLNA			
PWGARDNYGKKRRSRKXHQESTTDSLSHY5		FMLTWKEMRAKYTAECTILKESAINIOLGRWHALSREOAKYVELARKERQLHMQLY			
731. .751		PWGARDNYGKKRRSRKXHQESTTDSLSHY5			
mat_peptide		LPGEGRCPSVPDDSDALGCPGSPAFQDSFSYHLLPRFTTELTSPSERILHPQVVSPL			
/gene="TCF-1"		LSASQPGQPHRPPAAPCAHRYSNRNLDRWPSNRHTFGRLOEPTP"			
/product="alternative ORF specific for TCF-ID"		803. .1141			
BASE COUNT	272 a 407 c 265 g 171 t	/gene="TCF-1"			
ORIGIN	/product="alternative ORF specific for TCF-ID"				
Query Match 91.1%; Score 16.4; DB 9; Length 1115;					
Best Local Similarity 94.4%; Pred. No. 4.3e+02;					
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	1 ccgagtggtgggtgtaat 18	mat_peptide			
Db		BASE COUNT 279 a 422 c 286 g 178 t			
	62 CCGAGGGGGTGGGTAAT 45	ORIGIN			
RESULT 11					
HSTCF1E/c	HSTCF1E 1165 bp mRNA PRI 09-JAN-1995				
LOCUS	H. sapiens TCF-1 mRNA for T cell factor 1 splice form E.				
DEFINITION	H. sapiens TCF-1 mRNA for T cell factor 1 (splice form A).				
ACCESSION	247362				
VERSION	247362.1 GI:619883				
KEYWORDS	splice form E; T cell factor 1.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.				
JOURNAL	The human high mobility group (HMG)-box transcription factor TCF-1:				
AUTHORS	novel isoforms due to alternative splicing and usage of a new exon				
TITLE	IXA				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1165)				
AUTHORS	van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D.,				
TITLE	Suljkerbuij, R., Geurts van Kessel, A. and Clevers, H.				
JOURNAL	The human T cell transcription factor-1 gene. Structure,				
AUTHORS	localization, and promoter characterization				
TITLE	J. Biol. Chem. 267 (12), 8530-8536 (1992)				
JOURNAL	92335082				
MEDLINE	3 (bases 1 to 1165)				
REFERENCE	Ballhausen, W.G.				
AUTHORS	Direct Submission				
TITLE	Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer				
JOURNAL	Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,				
FEATURES	Schwabachanlage 10, Erlangen, Germany, D-91054				
SOURCE	Location/Qualifiers				
mRNA	1. .1165	/organism="Homo sapiens"			
	/db_xref="taxon:9606"	/db_xref="taxon:9606"			
gene	/tissue_type="lymphoma"	/cell_type="T lymphocyte"			
	/cell_line="Jurkat"	/clone_lib="cDNA"			
CDS	1. .1165	/clone="FTCF-1a"			
	/gene="TCF-1"	/gene="TCF-1"			
/standard_name="T cell factor 1 splice form E"		/evidence=experimental			
/citation=[1]		1. .1254			
80. .889		/gene="TCF-1"			
See also		X59869-X59871.			
Location/Qualifiers		1. .1254			
/organism="Homo sapiens"		/db_xref="taxon:9606"			
/cell_type="T lymphocyte"		/cell_line="Jurkat"			
/clone_lib="cDNA"		/clone="FTCF-1a"			
1. .1254		/gene="TCF-1"			
/evidence=experimental		1. .1254			
/gene="TCF-1"		80. .889			


```
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1, splice form A"
/protein_id="CAA42526.1"
/db_xref="GI:36786"
/db_xref="SWISS-PROT:P36402"
/translation="MYKETVYSAFNLLMHYPPPSGAGOHPOPPHUKANOPPHGVPO
LSLYEHFNSPHTPPADISQKVHRPLOTDLISGFYSLTSGSMGOLPHTVSWFTHPS
LSLYEHFNSPHTPPADISQKVHRPLOTDLISGFYSLTSGSMGOLPHTVSWFTHPS
LMLSGVPGHPAAIPHPAIVPPSGKQLOEPDRNLKTOAESKAEREAKKPTIKKPLNA
FMYLKMRAKVIATCKLKESAAINQILGRWHAISREEQAKYIELARKERQLHMQLY
PWSARDNYGKKRRSRKREKHQESTTETNPRLKDGNSLSMSSSSPA"
misc_feature
539..769
/gene="TCF-1"
/feature="HMG box"
misc_feature
810
/gene="TCF-1"
/feature="alternative splice site"
BASE COUNT      320 a   391 c   335 g   208 t
ORIGIN

Query Match      91.1%; Score 16.4; DB 9; Length 1254;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
||||| |||||||
Db 140 CCGAGGGGGTGGGTAAT 123

RESULT 13
HSTCF1B/c
LOCUS      HSTCF1B      2814 bp      mRNA      PRI      17-JUN-1991
DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form B).
ACCESSION  X59870.1 X55329
VERSION    .X59870.1 GI:36787
KEYWORDS   DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2814)
AUTHORS   van de Wetering, M., Oosterwegel, M., Doijes, D. and Clevers, H.
TITLE     Identification and cloning of TCF-1, a T lymphocyte-specific
JOURNAL   EMBO J. 10 (1), 123-132 (1991)
REFERENCE  2 (bases 1 to 2814)
AUTHORS   van de Wetering, M., Oosterwegel, M., Doijes, D. and Clevers, H.
TITLE     Identification and cloning of TCF-1, a T lymphocyte-specific
JOURNAL   EMBO J. 10 (1), 123-132 (1991)
COMMENT   See also X59869-X59871.
FEATURES   Location/Qualifiers
            source
            1..2814
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="T lymphocyte"
            /cell_line="Jurkat and HPB-ALL"
            /clone_lib="CDNA"
            1..2814
            /gene="TCF-1"
            /evidence=experimental
            1..2814
            /gene="TCF-1"
            80..889
            /gene="TCF-1"
            /codon_start=1
            /product="T cell factor 1 splice form B"
            /protein_id="CAA42527.1"
            /db_xref="GI:36788"
            /db_xref="SWISS-PROT:P36402"
```

```
/translation="MYKETVYSAFNLLMHYPPPSGAGOHPOPPHUKANOPPHGVPO
LSLYEHFNSPHTPPADISQKVHRPLOTDLISGFYSLTSGSMGOLPHTVSWFTHPS
LMLSGVPGHPAAIPHPAIVPPSGKQLOEPDRNLKTOAESKAEREAKKPTIKKPLNA
FMYLKMRAKVIATCKLKESAAINQILGRWHAISREEQAKYIELARKERQLHMQLY
PWSARDNYGKKRRSRKREKHQESTTGGRRNAFGTYPKAAAPAPFLPMTVL"
misc_feature
539..769
/gene="TCF-1"
/feature="HMG box"
misc_feature
810
/gene="TCF-1"
/feature="alternative splice site"
BASE COUNT      763 a   831 c   648 g   572 t
ORIGIN

Query Match      91.1%; Score 16.4; DB 9; Length 2814;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
||||| |||||||
Db 140 CCGAGGGGGTGGGTAAT 123

RESULT 14
HSTCF1G/c
LOCUS      HSTCF1G      2855 bp      DNA      PRI      30-SEP-1999
DEFINITION Homo sapiens TCF-1 gene.
ACCESSION  X63901
VERSION    X63901.1 GI:36791
KEYWORDS   T-cell transcription factor; transcription factor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2855)
AUTHORS   van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D.,
            Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.
TITLE     The human T cell transcription factor-1 gene. Structure,
JOURNAL   J. Biol. Chem. 267 (12), 8530-8536 (1992)
REFERENCE  2 (bases 1 to 2855)
AUTHORS   Van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H.
TITLE     Extensive alternative splicing and dual promoter usage generate
JOURNAL   Mol. Cell. Biol. 16 (3), 745-752 (1996)
REFERENCE  3 (bases 1 to 2855)
AUTHORS   Van de Wetering, M.L.
TITLE     Direct Submission
JOURNAL   Submitted (07-JAN-1992) M.L. Van De Wetering, Department of
            Immunology, P.O. Box 8550, 3508 GA Utrecht, THE NETHERLANDS
FEATURES   Location/Qualifiers
            1..2855
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /cell_type="T-lymphocyte"
            /clone="phi-TCF-[5,8,11]"
            /map="5q31.1"
            829..924
            /gene="TCF-1"
            /number=1
            join(829..924,970..1074,1120..1209,1255..1374,1420..1581,
            1691..1798,1843..1890,2054..2134)
            /gene="TCF-1"
            /codon_start=1
            /protein_id="CA856795.1"
            /db_xref="GI:600656"
            /translation="MYKETVYSAFNLLMHYPPPSGAGOHPOPPHUKANOPPHGVPO
            LSLYEHFNSPHTPPADISQKVHRPLOTDLISGFYSLTSGSMGOLPHTVSWFTHPS"
```

LMLGSGVGHGHPAAIPHPATVPSPGKQELQDFDRNLKTOAESKAEKAKPTIKKPLNA
FMLYKEMRAKVIAECTLKESAAINQILGRWHALSREEQAKYYELARKERQLHMQLY
PGWSARDNYGKKRRSRKREKHQESTTETNNPRLKDGNGQESLSMSSSSSPA"
join(829..924,970..1074,1120..1209,1255..1374,1420..1581,
1691..1798,1843..1890,2054..2134)
/gene="TCF-1"
925..970
/number=1
970..1074
/gene="TCF-1"
/number=2
1075..1119
/number=2
1120..1209
/gene="TCF-1"
/number=3
1210..1254
/number=3
1255..1374
/gene="TCF-1"
/number=4
1375..1419
/number=4
1420..1581
/gene="TCF-1"
/number=5
1582..1690
/number=5
1691..1798
/gene="TCF-1"
/number=6
1799..1842
/number=6
1843..1890
/gene="TCF-1"
/number=7
1891..2053
/number=1
2054..2134
/gene="TCF-1"
/number=8

BASE COUNT 592 a 929 c 771 g 562 t 1 others

Query Match 91.1%; Score 16.4; DB 9; Length 2855;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaqgtgggtggtgtaat 18
||||| |||||||||
Db 889 CCGAGGGGGTGGGTAAT 872

RESULT 15
HSTCF1C/H HSTCF1C 2910 bp mRNA PRI 17-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form C).
DEFINITION X59871 X55328
ACCESSION X59871.1 GI:36789
VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2910)
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
REFERENCE 2 (bases 1 to 2910)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H.

TITLE Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box
JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 91114695
COMMENT See also X59869-X59871.
FEATURES
Location/Qualifiers
1..2910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat and HPB-ALL"
/clone_lib="cDNA"
1..2910
/gene="TCF-1"
/evidence="experimental"
1..2910
/gene="TCF-1"
80..886
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1, splice form C"
/protein_id="CAA42528.1"
/db_xref="GI:36790"
/db_xref="SWISS-PROT:P36402"
/translation="MYKETVYSAFNLLMHYPPSPGAGOHPOPPPLHKANOPPHGVPO
LSLYEHFNSPHPTAPADISQKQVHRPLQTPDLSGFYSLTSGMQLPHTYSWTFHPS
LMLGSGVGHGHPAAIPHPATVPSPGKQELQDFDRNLKTOAESKAEKAKPTIKKPLNA
FMLYKEMRAKVIAECTLKESAAINQILGRWHALSREEQAKYYELARKERQLHMQLY
PGWSARDNYGKKRRSRKREKHQESTTDPGSPKKCRARFGLNQOUTWCGPCR"
539..769
/gene="TCF-1"
/note="HMG box"
810
/gene="TCF-1"
/note="alternative splice site"

BASE COUNT 782 a 860 c 671 g 597 t

Query Match 91.1%; Score 16.4; DB 9; Length 2910;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaqgtgggtggtgtaat 18
||||| |||||||||
Db 140 CCGAGGGGGTGGGTAAT 123

Search completed: February 20, 2002, 11:03:22
Job time: 12110 sec

c 12	15	83.3	1072	20	AA57478	Rat U3 gene trap d
c 13	15	83.3	596	20	AAV83941	Bacterial artifici
c 14	15	83.3	80240	20	AAV83940	NC-contig derived
c 15	14.8	82.2	115	22	AAI63452	Human kidney relat
c 16	14.8	82.2	128	22	AAI63451	Human kidney relat
c 17	14.8	82.2	132	21	AAI14418	Human secreted pro
c 18	14.8	82.2	138	21	AAI13252	Human secreted pro
c 19	14.8	82.2	138	21	AAI15130	Human secreted pro
c 20	14.8	82.2	168	22	AAI63460	Human kidney relat
c 21	14.8	82.2	169	22	AAI62840	Human genomic DNA
c 22	14.8	82.2	240	21	AAI13781	Human secreted pro
c 23	14.8	82.2	312	21	AAI21674	Human secreted pro
c 24	14.8	82.2	343	18	AAI51191	Human breast speci
c 25	14.8	82.2	392	14	AAO61201	Human brain Expre
c 26	14.8	82.2	404	21	AAO04048	Human secreted pro
c 27	14.8	82.2	435	21	AAI17237	Human secreted pro
c 28	14.8	82.2	469	22	AAI64073	Human bladder rela
c 29	14.8	82.2	543	22	AAH09911	Human cDNA clone (
c 30	14.8	82.2	559	22	AAH09755	Human cDNA clone (
c 31	14.8	82.2	560	22	AAH09448	Human cDNA clone (
c 32	14.8	82.2	582	22	AAH09374	Human cDNA clone (
c 33	14.8	82.2	700	22	AAH92769	Human inflammatory
c 34	14.8	82.2	700	22	AAH92770	Human inflammatory
c 35	14.8	82.2	704	21	AAAO1949	Human colon cancer
c 36	14.8	82.2	777	22	AAH03919	Human cDNA clone (
c 37	14.8	82.2	909	22	AAH19222	Human secreted pro
c 38	14.8	82.2	998	20	AAV27362	Human secreted pro
c 39	14.8	82.2	999	22	AAAF84751	Nucleotide sequenc
c 40	14.8	82.2	1272	21	AAAC44779	Arabidopsis thalia
c 41	14.8	82.2	1344	21	AAAC44779	Zea mays DNA fragm
c 42	14.8	82.2	1432	22	AAAO5450	Human secreted pro
c 43	14.8	82.2	1647	22	AAH17150	Human cDNA sequenc
c 44	14.8	82.2	1648	22	AAH13852	Human cDNA sequenc
c 45	14.8	82.2	1654	21	AAAC59710	Human secreted pro

ALIGNMENTS

RESULT 1						
AAC14388/c						
ID	AAC14388	standard;	cDNA; 227 BP.			
XX	AAC14388;					
AC	AAC14388;					
DT	06-OCT-2000	(first entry)				
XX	Human secreted protein 5' EST, SEQ ID NO: 18463.					
DE	Human secreted protein 5' EST, SEQ ID NO: 18463.					
XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;					
KW	gene therapy; chromosome mapping; ss.					
XX	Homo sapiens.					
OS	Homo sapiens.					
XX	EPI033401-A2.					
XX	PD					
PD	06-SEP-2000.					
XX	21-FEB-2000; 2000EP-0200610.					
XX	26-FEB-1999; 99US-0122487.					
PR	(GEST) GENSET.					
PA	Dumas Milne Edwards J, Duclert A, Giordano J;					
XX	WPI: 2000-500381/45.					
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for					
PT	obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for					
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -					
XX	Claim 1; SEQ ID 18463; 71pp + CD-ROM; English.					
PS						

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 20, 2002, 10:32:01 ; Search time 363.89 Seconds
(without alignments)
42.408 Million cell updates/sec

Title: US-09-904-420A-9
Perfect score: 18
Sequence: 1 ccgagtggtgggtaat 18
Scoring table: IDENTITY_NUC
Gapop 10.0, capext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_1101.*
1:	/SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	15.4	85.6	227	AAI4388	Human secreted pro
c 2	15.4	85.6	244	AAI11782	Human secreted pro
c 3	15.4	85.6	776	AAI93454	Human secreted pro
c 4	15.4	85.6	1604	AAH18451	Human cDNA sequenc
c 5	15.4	85.6	4372	AAH18451	Human cDNA sequenc
c 6	15.4	85.6	9365	AAI20359	Human CD39-like pr
c 7	15.4	85.6	9365	AAI20359	Human CD39-like pr
c 8	15.4	85.6	14747	AAAF63405	Human CD39-like pr
c 9	15.4	85.6	15977	AAAF63406	Human breast or ov
c 10	15.4	85.6	16225	AAI62650	Human yes1 gene.
c 11	15.4	85.6	119950	AAI90201	

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 227 BP; 44 A; 77 C; 53 G; 53 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 227;
 Best Local Similarity 94.1%; Pred. No. 97;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaggtgggtgggttaa 17
 |||||
 Db 184 CCGAGGTGGTGGGTCA 168

RESULT 2

AAC11782
 ID AAC11782 standard; cDNA; 244 BP.
 XX
 AC AAC11782;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 15857.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PP 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI; 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 15857; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 244 BP; 62 A; 57 C; 59 G; 66 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 244;
 Best Local Similarity 94.1%; Pred. No. 98;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaggtgggtgggttaa 17
 |||||
 Db 104 ccgaggtgggtgggtca 120

RESULT 3

AAC93454
 ID AAC93454 standard; cDNA; 776 BP.
 XX
 AC AAC93454;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein gene 33 SEQ ID NO:43.
 XX
 KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neutropic; neuroprotective; antibacterial; virucide; fungicide;
 KW hypermalignant; vulnary; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2000061625-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US08981.
 XX
 PR 09-APR-1999; 99US-0128701.
 PR 20-JAN-2000; 2000US-0177166.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI; 2000-619226/59.
 DR P-PSDB; AAB51762.
 XX
 PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; Page 435; 500pp; English.

XX Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding
 CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -
 CC AAB51825 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences to which they are homologous. The genes and proteins
 CC have activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;
 CC virucide; fungicide; opthalmological; and vulnary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. Oligonucleotide AAC93413 - AAC93421 and peptide AA851723 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention.
 XX
 SQ Sequence 776 BP; 211 A; 171 C; 201 G; 193 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 776;
 Best Local Similarity 94.1%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaggtgggtgggtaa 17
 |||||

Db 519 ccgaggtgggtgggtaa 535

RESULT 4
 AAH18451
 ID AAH18451 standard; cDNA; 1604 BP.
 AC AAH18451;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18547.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 8; SEQ ID 18547; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1604 BP; 413 A; 339 C; 376 G; 476 T; 0 other;

Query Match 85.6%; Score 15.4; DB 22; Length 1604;
 Best Local Similarity 94.1%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgaggtgggtgggtaa 17
 |||||

Db 585 ccgaggtgggtgggtaa 601

RESULT 5
 AAH18649
 ID AAH18649 standard; cDNA; 4372 BP.
 AC AAH18649;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SFQ ID NO:18880.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 8; SEQ ID 18880; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 4372 BP; 1270 A; 862 C; 1032 G; 1208 T; 0 other;

Query Match 85.6%; Score 15.4; DB 22; Length 4372;
 Best Local Similarity 94.1%; Pred. No. 1.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggttaa 17
 |||||
 Db 3127 ccgagtggtgggttaa 3143

RESULT 6
 AAZ50359
 ID AAZ50359 standard; DNA; 9365 BP.

XX AC AAZ50359;

XX DT 18-MAY-2000 (first entry)

XX DE Human CD39-L4 genomic DNA.

XX KW CD39-L4; human; apyrase; nucleotide diphosphatase; NDPase;
 KW ATP diphosphohydrolase; ATPDase; adenosine diphosphate; ADP; treatment;
 KW platelet aggregation; antithrombotic; thrombosis; myocardial infarction;
 KW cerebral ischaemia; angina; vascular graft; extracorporeal circulation;
 KW molecular weight marker; nutritional supplement; tumour; prevention;
 KW drug targeting; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 72..8642

FT /tag= a
 FT /product= "Human CD39-L4 protein"
 FT /note= "Coding region is interrupted with introns"

FT FT exon 1..288

FT FT /tag= b

FT FT /number= 1

FT FT 289..1280

FT FT /tag= c

FT FT /number= 1

FT FT 1281..1580

FT FT /tag= d

FT FT /number= 2

FT FT 1581..1819

FT FT /tag= e

FT FT /number= 2

FT FT 1820..1855

FT FT /tag= f

FT FT /number= 3

FT FT 1856..2466

FT FT /tag= g

FT FT /number= 3

FT FT 2467..2555

FT FT /tag= h
 FT /number= 4
 FT 2556..2862
 FT /tag= i
 FT /number= 4
 FT 2863..2942
 FT /tag= j
 FT /number= 5
 FT 2943..3888
 FT /tag= k
 FT /number= 5
 FT 3889..3950
 FT /tag= l
 FT /number= 6
 FT 3951..4893
 FT /tag= m
 FT /number= 6
 FT 4894..4995
 FT /tag= n
 FT /number= 7
 FT 4996..5846
 FT /tag= o
 FT /number= 7
 FT 5847..5987
 FT /tag= p
 FT /number= 8
 FT 5988..6965
 FT /tag= q
 FT /number= 8
 FT 6966..7138
 FT /tag= r
 FT /number= 9
 FT 7139..8555
 FT /tag= s
 FT /number= 9
 FT 8556..9365
 FT /tag= t
 FT /number= 10

XX WO200004041-A2.

XX PN 27-JAN-2000.

XX PD 16-JUL-1999; 99WO-US16180.

XX PF 16-JUL-1998; 98US-0118205.

XX PR 24-JUL-1998; 98US-0122449.

XX PR 04-FEB-1999; 99US-0244444.

XX PR 19-MAR-1999; 99US-0273447.

XX PR 09-JUL-1999; 99US-0350836.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero J;

XX WPI: 2000-182397/16.

XX P-PSDB; AAY44849.

XX New nucleic acid encoding human CD39-like protein, useful for treating
 XX and preventing thrombotic disease -

XX Example 11; Page 112-119; 125pp; English.

XX The present sequence is the genomic DNA encoding CD39-L4 protein, an
 XX apyrase and/or nucleotide diphosphatase (NDPase). It is isolated from
 XX the human C17B BAC genomic library. It is a soluble ATP
 XX Diphosphohydrolase (ATPDase) and is involved in the hydrolysis of
 XX adenosine diphosphate (ADP), the agonist that causes platelet
 XX aggregation. CD39-L4 protein has 30% and 80% homology to human and
 XX murine CD39. It has platelet aggregation inhibition and antithrombotic
 XX activity. CD39-L4 is used to treat or prevent thrombosis, myocardial
 XX infarction, cerebral ischaemia and angina. It is also used in vitro, to
 XX maintain vascular grafts or during extracorporeal circulation, to

CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
CC modulating disease states (including platelet aggregation, inflammation
CC and apoptosis) associated with ADP or other purinergic signalling by
CC reducing the levels of NDPs. The polypeptides are also useful for
CC prophylaxis or treatment of inflammation related disorders, such as
CC disorders involving sepsis or systemic inflammatory response syndrome or
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
CC cytokine overstimulation); autoimmune disorders such as thrombosis,
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
CC neurological disorders including neurodegenerative diseases, epilepsy,
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, and amyotrophic lateral sclerosis; and cancer. The present
CC sequence represents the CD39 like protein CD39-L4 genomic DNA sequence.
XX
SQ Sequence 14747 BP; 3821 A; 3235 C; 3349 G; 4294 T; 48 other;

Query Match 85.6%; Score 15.4; DB 22; Length 14747;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaa 17
Db 6081 ccgaggtgggtgggtca 6097

RESULT 9
AAF63407
ID AAF63407 standard; DNA; 15977 BP.

XX AAF63407;

XX 14-MAY-2001 (first entry)

XX Human CD39 like protein CD39-L4 genomic DNA sequence #2.

XX Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;
KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
KW cerebral artery thrombosis; platelet aggregation; inflammation;
KW apoptosis; autoimmune disorder; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L4; ds.

OS Homo sapiens.

XX WO200110205-A1.

XX 15-FEB-2001.

XX 09-AUG-2000; 2000WO-US21790.

XX 09-AUG-1999; 99US-0370265.

PR 11-JAN-2000; 2000US-0481238.

PR 25-APR-2000; 2000US-0557800.

PR 26-MAY-2000; 2000US-0583231.

PR 30-JUN-2000; 2000US-0608285.

XX (HYSE-) HYSEQ INC.

XX Ford J, Mulero JJ, Yeung G;

XX WPI; 2001-147489/15.

XX Polynucleotides encoding human CD39-like polypeptides, with apyrase
XX and/or NDPase activity, which are useful in the treatment of
PT pathological conditions caused by thrombosis (e.g. myocardial
PT infarction) and inflammatory disorders -

XX Example 11: Page 187-192; 203pp; English.

XX This invention relates to polynucleotides encoding human CD39-like
XX polypeptides with apyrase and/or NDPase activity. The polypeptides having
CC ATPase, including NDPase, activity are useful for inhibiting platelet

CC function and can therefore be used in the prophylaxis or treatment of
CC pathological conditions caused by or involving thrombosis or excessive
CC coagulation or excessive platelet aggregation, such as myocardial
CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral
CC artery thrombosis or intracardiac thrombosis, and conditions associated
CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
CC modulating disease states (including platelet aggregation, inflammation
CC and apoptosis) associated with ADP or other purinergic signalling by
CC reducing the levels of NDPs. The polypeptides are also useful for
CC prophylaxis or treatment of inflammation related disorders, such as
CC disorders involving sepsis or systemic inflammatory response syndrome or
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
CC cytokine overstimulation); autoimmune disorders such as thrombosis,
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
CC neurological disorders including neurodegenerative diseases, epilepsy,
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, and amyotrophic lateral sclerosis; and cancer. The present
CC sequence represents the CD39 like protein CD39-L4 genomic DNA sequence.
XX
SQ Sequence 15977 BP; 4202 A; 3489 C; 3648 G; 4592 T; 46 other;

Query Match 85.6%; Score 15.4; DB 22; Length 15977;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaa 17
Db 6081 ccgaggtgggtgggtca 6097

RESULT 10

AAF62650/C

ID AAF62650 standard; DNA; 16225 BP.

XX AAF62650;

XX 19-OCT-2001 (first entry)

XX Human breast or ovarian antigen genomic DNA SEQ ID NO: 300.

XX Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
KW ds.

OS Homo sapiens.

XX WO2001155324-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01344.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

us-09-904-420a-9.rng

Thu Feb 21 07:26:08 2002

```
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488785/53.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
XX treating and/or preventing human diseases and disorders -
XX
XX Disclosure; SEQ ID NO: 300; 520pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of ovarian and breast antigens. These are shown in
XX AA162467-AA162572 and AAM42240-AAM42345. The sequences can be used in the
XX diagnosis, prevention and treatment of breast and ovarian cancers, and
XX their metastases. The present sequence is a genomic sequence of the
XX invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 16225 BP; 4434 A; 3583 C; 3702 G; 4506 T; 0 other;
XX
SQ
Query Match 85.6%; Score 15.4; DB 22; Length 16225;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ccgaggtgggtgggtaa 17
```

Thu Feb 21 07:26:08 2002

Db

8511

CCGAGTGGTGGATAA 8495

RESULT 11

AA90201

ID AAX90201 standard; DNA; 119950 BP.

XX

AC AAX90201;

DT 23-SEP-1999 (first entry)

XX

DE Human yes1 gene.

XX

KW Human; yes1; diagnosis; neuropsychiatric disorder; BAD: schizophrenia;

KW bipolar affective disorder; attention deficit disorder;

KW schizoaffective disorder; unipolar affective disorder;

XX

OS Huntington's disease; Parkinson's disease; manic-depression; ds.

XX

PN Homo sapiens.

XX

PN WO9935290-A1.

XX

PD 15-JUL-1999.

XX

PF 07-JAN-1999; 99WO-US00297.

XX

PR 08-JAN-1998; 98US-0003944.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Chen H, Freimer NB;

XX

DR WPI: 1999-444203/37.

XX

DR P-PSDB; AAY24421.

XX

PT Detection of a genetic mutation in the yes1 gene, useful for

PT diagnosis of a yes1 mediated neuropsychiatric disorder in a human

XX

XX Claim 1; Fig 2; 110pp; English.

XX

PS The present invention describes a method for detecting a genetic

CC mutation in the yes1 gene for the diagnosis of a yes1 mediated

CC neuropsychiatric disorder in a human. The method comprises detecting the

CC presence or absence of a genetic mutation in the yes1 gene of the

CC subject, where the genetic mutation is a substitution, insertion or a

CC deletion and results in the production of a yes1 protein having an amino

CC acid sequence other than the wild-type yes1 amino acid sequence and the

CC presence of the genetic mutation identifies a subject that has or is at

CC risk for developing a yes1 mediated neuropsychiatric disorder. Compounds

CC that bind to the yes1 protein, alter the amount of the protein, or alter

CC the activity of the yes1 gene product, are useful for treating a yes1

CC mediated neuropsychiatric disorder. The disorders include Huntington's

CC disease, Parkinson's disease, and especially bipolar-affective disorder

CC (BAD) also known as bipolar mood disorder (BP) or manic-depressive

CC illness. The method distinguishes neuropsychiatric disorders from

CC neurological disorders, which enables more accurate evaluation and

CC prescription of medical treatment. The present sequence represents the

CC human yes1 cDNA sequence.

XX

XX Sequence 119950 BP; 34471 A; 23730 C; 24660 G; 37033 T; 56 other;

Query Match 85.6%; Score 15.4; DB 20; Length 119950;

Best Local Similarity 94.1%; Pred. No. 1.2e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ccgaggtgggtgggttaa 17

Db 94111 ccgaggtgggtgggtcaa 94127

|||||

AAX57478/c

ID AAX57478 standard; DNA; 1072 BP.

XX

AC AAX57478;

DT 24-JUL-1999 (first entry)

XX

DE Rat U3 gene trap derived nucleic acid 31_4_2-re.

XX

KW Gene trap; rat; cellular gene; viral infection; cell survival; cancer;

KW tumour progression; suppression; identification; viral growth;

KW tumour suppressor; prevention; screening; therapeutic agent; ss.

XX

OS Rattus norvegicus.

XX

PN WO9919481-A2.

XX

PD 22-APR-1999.

XX

PF 08-OCT-1998; 98WO-US21276.

XX

PR 10-OCT-1997; 97US-0062021.

XX

XX (UYVA-) UNIV VANDERBILT.

XX

PI Dubois RN, Organ EL, Rubin DH;

XX

DR WPI: 1999-326546/27.

XX

PT Nucleic acid encoding tumor suppressors and products required for

PT viral infection

XX

PS Claim 1; Page 85; 94pp; English.

XX

CC This invention describes novel rat-derived nucleic acid fragments from

CC cellular genes that are necessary for viral infection but not for cell

CC survival, or that suppress tumour progression. The products of the

CC invention (AAX57371-X57497) can be used in methods of identifying

CC cellular genes necessary for viral growth and cellular genes that

CC function as tumour suppressors and for reducing or preventing such

CC infections or cancer. They may also be used in screening for potential

CC therapeutic agents. These sequences can be targeted without significant

CC side effects (contrast targeting genes essential for viral growth).

XX

XX Sequence 1072 BP; 251 A; 235 C; 298 G; 241 T; 47 other;

Query Match 83.3%; Score 15; DB 20; Length 1072;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gaggtgggtgggttaa 17

Db 740 GAGGTGGTGGGTAA 726

|||||

RESULT 13

AAV83941/c

ID AAV83941 standard; DNA; 5596 BP.

XX

AC AAV83941;

XX

DT 03-MAR-1999 (first entry)

XX

DE Bacterial artificial chromosome (BAC)-F2 contig 1.

XX

KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;

KW neocentromere; replication; extra-chromosomal element; segregation;

KW cell division; artificial chromosome; gene therapy; BAC; transgenic;

XX

OS human artificial chromosome; bacterial artificial chromosome; ss.

XX

OS Synthetic.

PN WO9851790-A1.
 XX 19-NOV-1998.
 XX 13-MAY-1998; 98WO-AU00352.
 XX 26-AUG-1997; 97AU-0008791.
 PR 13-MAY-1997; 97AU-0006784.
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 PA Cancilla MR, Choo K, Du Sart D;
 XX WPI; 1999-009773/01.
 XX New isolated nucleic acid comprising neocentromere sequences from
 PT eukaryotic chromosome - used to produce replicable, segregating
 PT artificial chromosomes that can carry large amounts of DNA for gene
 PT therapy
 XX Claim 10; Page 177-181; 540pp; English.
 XX The present sequence represents a bacterial artificial chromosome (BAC)
 CC contig, and exemplifies the invention. The specification describes
 CC nucleic acid sequences derived from a eukaryotic chromosome, including a
 CC neocentromere or its functional derivative or hybrid, that are able, in
 CC a compatible cell, of replicating, acting as extra-chromosomal element
 CC and segregating during cell division. The sequences can be used to
 CC construct artificial chromosomes for use in gene therapy comprising a
 CC replicable, segregating nucleic acid that confers a specific phenotype
 CC on cells. Human artificial chromosomes can propagate in human cells and
 CC carry large amounts of DNA (e.g. therapeutic genes), and, being
 CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
 CC are also useful for generation of transgenic plants and animals, in
 CC production of proteins and to make diagnostic reagents, e.g. for
 CC expression of cytokines, receptors and growth factors, or to increase
 CC the copy number of a gene in a cell. The constructs may also be
 CC used for functional and structural analysis of chromosomes.
 XX SQ Sequence 5596 BP; 1730 A; 1019 C; 1082 G; 1763 T; 2 other;

Query Match 83.3%; Score 15; DB 20; Length 5596;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aggtgggtgggtaat 18
 |||||
 Db 2992 AGGTGGTGGGTAAT 2978

RESULT 14
 AAV83940/C
 ID AAV83940 standard; DNA; 80240 BP.
 XX AAV83940;
 XX 03-MAR-1999 (first entry)
 XX NC-contig derived from mardel(10) on chromosome 10q25.2.
 XX Yeast artificial chromosome; YAC; probe: eukaryotic chromosome;
 KW neocentromere; replication; extra-chromosomal element; segregation;
 KW cell division; artificial chromosome; gene therapy; mardel(10);
 KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
 XX Homo sapiens.
 OS WO9851790-A1.
 PN 19-NOV-1998.
 XX 13-MAY-1998; 98WO-AU00352.

XX 26-AUG-1997; 97AU-0008791.
 PR 13-MAY-1997; 97AU-0006784.
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX Cancilla MR, Choo K, Du Sart D;
 XX WPI; 1999-009773/01.
 XX New isolated nucleic acid comprising neocentromere sequences from
 PT eukaryotic chromosome - used to produce replicable, segregating
 PT artificial chromosomes that can carry large amounts of DNA for gene
 PT therapy
 XX Claim 9; Fig 16A; 540pp; English.
 XX The present sequence represents the NC-contig derived from a mutated
 CC human chromosome 10, 10q25.2 region. The sequence contains
 CC an unusual chromosomal marker referred to as mardel(10). The
 CC mardel(10) marker is mitotically stable and contains a functional
 CC neocentromere at a location regarded as non-centromeric. This
 CC neocentromere maps to q25.2 on chromosome 10. The specification describes
 CC nucleic acid sequences derived from a eukaryotic chromosome, including a
 CC neocentromere or its functional derivative or hybrid, that are able, in
 CC a compatible cell, of replicating, acting as extra-chromosomal element
 CC and segregating during cell division. The sequences can be used to
 CC construct artificial chromosomes for use in gene therapy comprising a
 CC replicable, segregating nucleic acid that confers a specific phenotype
 CC on cells. Human artificial chromosomes can propagate in human cells and
 CC carry large amounts of DNA (e.g. therapeutic genes), and, being
 CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
 CC are also useful for generation of transgenic plants and animals, in
 CC production of proteins and to make diagnostic reagents, e.g. for
 CC expression of cytokines, receptors and growth factors, or to increase
 CC the copy number of a gene in a cell. The constructs may also be
 CC used for functional and structural analysis of chromosomes.
 XX SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 8 other;

Query Match 83.3%; Score 15; DB 20; Length 80240;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aggtgggtgggtaat 18
 |||||
 Db 70981 AGGTGGTGGGTAAT 70967

RESULT 15
 AAI63452
 ID AAI63452 standard; DNA; 115 BP.
 XX AAI63452;
 XX 22-OCT-2001 (first entry)
 XX Human kidney related polynucleotide SEQ ID NO 767.
 XX Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neotropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
 KW antinflammatory; antitumor; cancer; immune disorder; cardiovascular disorder;
 KW gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; ds.
 XX Homo sapiens.
 OS WO200155323-A2.
 PN 02-AUG-2001.

XX	17-JAN-2001; 2001WO-US01343.	PR	29-SEP-2000; 2000US-0236369.
PF		PR	29-SEP-2000; 2000US-0236370.
XX	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0236802.
XX	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237037.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237038.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000; 2000US-0237039.
PR	16-MAR-2000; 2000US-0189874.	PR	02-OCT-2000; 2000US-0237040.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000; 2000US-0239935.
PR	18-APR-2000; 2000US-0198123.	PR	13-OCT-2000; 2000US-0239937.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0240960.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241221.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241785.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241786.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216680.	PR	20-OCT-2000; 2000US-0241808.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217496.	PR	01-NOV-2000; 2000US-0241826.
PR	14-JUL-2000; 2000US-0218290.	PR	01-NOV-2000; 2000US-0244617.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246475.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224521.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246610.
PR	18-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246611.
PR	22-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226688.	PR	17-NOV-2000; 2000US-0249208.
PR	23-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249209.
PR	30-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249210.
PR	01-SEP-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249214.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249215.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249216.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249217.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249218.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000; 2000US-0249299.
PR	12-SEP-2000; 2000US-0232081.	PR	17-NOV-2000; 2000US-0249300.
PR	14-SEP-2000; 2000US-0231968.	PR	01-DEC-2000; 2000US-0250160.
PR	14-SEP-2000; 2000US-0232397.	PR	01-DEC-2000; 2000US-0250391.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000; 2000US-0251030.
PR	14-SEP-2000; 2000US-0232399.	PR	05-DEC-2000; 2000US-0251988.
PR	14-SEP-2000; 2000US-0232400.	PR	06-DEC-2000; 2000US-0256719.
PR	14-SEP-2000; 2000US-0232401.	PR	08-DEC-2000; 2000US-0251479.
PR	14-SEP-2000; 2000US-0233063.	PR	08-DEC-2000; 2000US-0251856.
PR	14-SEP-2000; 2000US-0233064.	PR	08-DEC-2000; 2000US-0251868.
PR	14-SEP-2000; 2000US-0233065.	PR	08-DEC-2000; 2000US-0251869.
PR	21-SEP-2000; 2000US-0234223.	PR	08-DEC-2000; 2000US-0251989.
PR	21-SEP-2000; 2000US-0234274.	PR	11-DEC-2000; 2000US-0251990.
PR	25-SEP-2000; 2000US-0234997.	PR	05-JAN-2001; 2001US-0254097.
PR	25-SEP-2000; 2000US-0234998.	XX	
PR	26-SEP-2000; 2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.
PR	27-SEP-2000; 2000US-0235834.	XX	
PR	27-SEP-2000; 2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;
PR	29-SEP-2000; 2000US-0236327.	XX	
PR	29-SEP-2000; 2000US-0236367.	DR	WPI; 2001-488784/53.
PR	29-SEP-2000; 2000US-0236368.	XX	

PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Disclosure: SEQ ID NO 767; 564pp + Sequence Listing: English.
XX
CC The invention relates to novel kidney related polynucleotides
CC (AAI62971-AAI63793) and the encoded polypeptides (AAM42417-AAM42691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polynucleotides and proteins are also
CC useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. The genes are isolated from a range
CC of human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections
CC
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 115 BP; 28 A; 27 C; 35 G; 25 T; 0 other;

Query Match 82.2%; Score 14.8; DB 22; Length 115;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
| | | | | | | | | | | | | | | | | |
Db 43 ccgaggtgggtggatcat 60

Search completed: February 20, 2002, 10:32:12
Job time: 10240 sec

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:44:50 ; Search time 166.82 seconds
(without alignments)
24.437 Million cell updates/sec

Title: US-09-904-420A-9
Perfect score: 18
Sequence: 1 ccgaggtgggtgggtaat 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS-COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
C 1	15.4	85.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C 2	15.4	85.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
C 3	15	83.3	5596	4	US-09-078-294-5	Sequence 5, Appli
C 4	15	83.3	80246	4	US-09-078-294-4	Sequence 4, Appli
C 5	14.8	82.2	343	5	PCT-US95-08295-13	Sequence 13, Appli
C 6	14.8	82.2	1839	4	US-09-095-385-3	Sequence 3, Appli
C 7	14.8	82.2	2686	3	US-09-358-384-1	Sequence 1, Appli
C 8	14.8	82.2	2919	3	US-08-434-000A-3	Sequence 3, Appli
C 9	14.8	82.2	3224	4	US-09-312-157-3	Sequence 3, Appli
C 10	14.8	82.2	3224	4	US-08-965-729A-2	Sequence 2, Appli
C 11	14.8	82.2	3416	2	US-08-724-394A-15	Sequence 15, Appli
C 12	14.8	82.2	7210	2	US-08-257-963B-10	Sequence 10, Appli
C 13	14.8	82.2	7210	5	PCT-US95-07201-10	Sequence 10, Appli
C 14	14.8	82.2	22481	5	PCT-US95-07201-43	Sequence 43, Appli
C 15	14.8	82.2	36741	4	US-09-301-665-3	Sequence 3, Appli
C 16	14.8	82.2	49136	4	US-09-422-869-1	Sequence 1, Appli
C 17	14.4	80.0	1161	1	US-08-383-750-5	Sequence 5, Appli
C 18	14.4	80.0	1161	2	US-08-684-687-1	Sequence 1, Appli
C 19	14.4	80.0	1161	3	US-08-352-678-5	Sequence 5, Appli
C 20	14.4	80.0	1164	5	PCT-US93-09636-5	Sequence 5, Appli
C 21	14.4	80.0	1206	3	US-09-081-320-34	Sequence 34, Appli
C 22	14.4	80.0	1296	3	US-09-081-320-38	Sequence 38, Appli
C 23	14.4	80.0	6485	3	US-09-081-320-2	Sequence 2, Appli
C 24	14.4	80.0	8743	3	US-09-081-320-1	Sequence 1, Appli
C 25	14	77.8	12847	1	US-08-550-715-1	Sequence 1, Appli
C 26	13.8	76.7	282	1	US-08-133-629-8	Sequence 8, Appli
C 27	13.8	76.7	302	2	US-08-849-701-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 85.6%; Score 15.4; DB 4; Length 4403765;

Best Local Similarity 94.1%; Pred. No. 33;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 2 ccgagtggtgggtaat 18

|||||

Db 4325921 CGAGGTGGTGGGAAT 4325905

|||||

RESULT 2

US-09-103-840A-1/c

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 85.6%; Score 15.4; DB 4; Length 4411529;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 cgaggtgggtgggtaatt 18
|||||
DB 4333667 CGAGGTGGGTGGGTAAT 4333651

RESULT 3
US-09-078-294-5/c
; Sequence 5, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5596
; TYPE: DNA
; ORGANISM: BAC-F2 contig 1
US-09-078-294-5

Query Match 83.3%; Score 15; DB 4; Length 5596;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aggtgggtgggtaatt 18
|||||
DB 2992 AGGTGGGTGGGTAAT 2978

RESULT 4
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 83.3%; Score 15; DB 4; Length 80246;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aggtgggtgggtaatt 18
|||||
DB 70981 AGGTGGGTGGGTAAT 70967

RESULT 5
PCT-US95-08295-13
; Sequence 13, Application PC/TUS9508295
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: BREAST SPECIFIC GENES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08295
; FILING DATE: 30-JUN-1995
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-08295-13

Query Match 82.2%; Score 14.8; DB 5; Length 343;
Best Local Similarity 88.9%; Pred. No. 82;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccgaggtgggtgggtaatt 18
|||||
DB 52 CCCAGGAGGTGGGTAAT 69

RESULT 6
US-09-095-385-3/c
; Sequence 3, Application US/09095385
; Patent No. 6300104
; GENERAL INFORMATION:
; APPLICANT: Morrison, Sherie L.
; APPLICANT: Chintalacharuvu, Kote R.
; TITLE OF INVENTION: SECRETORY IMMUNOGLOBULIN PRODUCED
; TITLE OF INVENTION: BY SINGLE CELLS AND METHODS FOR MAKING AND USING
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,385
; FILING DATE: 09-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/050,969
; FILING DATE: 19-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Canady, Karen S

Thu Feb 21 07:26:09 2002

us-09-904-420a-9.rni

```

; REGISTRATION NUMBER: 39,927
; REFERENCE/DOCKET NUMBER: 30435.45USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-09-095-385-3

```

```

Query Match      82.2%; Score 14.8; DB 4; Length 1839;
Best Local Similarity 88.9%; Pred. No. 85;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ccgaggtgggtgggtaat 18
   | ||||| ||||| ||
DB 139 CAGAGTGGTGGGTAGT 122

```

RESULT 7

```

; US-09-358-384-1/c
; Sequence 1, Application US/09358384
; Patent No. 6130088
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF TELOMERIC REPEAT BINDING FACTOR 1 EXPRESSION
; FILE REFERENCE: RTS-0083
; CURRENT APPLICATION NUMBER: US/09/358,384
; CURRENT FILING DATE: 1999-07-21
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)..(1335)
; US-09-358-384-1

```

```

Query Match      82.2%; Score 14.8; DB 3; Length 2686;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ccgaggtgggtgggtaat 18
   | ||||| ||||| ||
DB 1940 CCGAGTGGTGGGTAGT 1923

```

RESULT 8

```

; US-08-434-000A-3/c
; Sequence 3, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.

```

```

; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELETYPE: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 235....2472
; US-08-434-000A-3

```

```

Query Match      82.2%; Score 14.8; DB 3; Length 2919;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ccgaggtgggtgggtaat 18
   | ||||| ||||| ||
DB 319 CAGAGTGGTGGGTAGT 302

```

RESULT 9

```

; US-09-312-157-3/c
; Sequence 3, Application US/09312157
; Patent No. 6303341
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/312,157

```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3224 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE::
; NAME/KEY: misc_feature
; OTHER INFORMATION: /note= "Nucleotides 2272 through 2729 are a
; OTHER INFORMATION: thrombin responsive element"; Human
US-08-965-729A-2

Query Match      82.2%; Score 14.8; DB 4; Length 3224;
Best Local Similarity 88.9%; Pred.No.86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; G

QY      1 ccgaggtgggtgggtaat 18
          |||||
Db       317 CCGAGGTGGTGCATCAT 300

RESULT 11
US-08-724-394A-15/c
; Sequence 15, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..3416
; OTHER INFORMATION: "CDNA 44"
```

US-08-724-394A-15

Query Match 82.2%; Score 14.8; DB 2; Length 3416;
 Best Local Similarity 88.9%; Pred. No. 86;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccgaggtgggtggtgtaatt 18
 |||||
 DB 3146 CCGAGGTGGTGGATCAT 3129

RESULT 12
 US-08-257-963B-10
 ; Sequence 10, Application US/08257963B
 ; Patent No. 5840686
 ; GENERAL INFORMATION:
 ; APPLICANT: Chader, Gerald J.; Becerra, S.
 ; APPLICANT: Patricia; Schwartz, Joan P.;
 ; APPLICANT: Taniwaki, Takayuki
 ; TITLE OF INVENTION: PIGMENT EPITHELIUM
 ; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
 ; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
 ; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morgan & Finnegan
 ; STREET: 345 Park Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/257,963B
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/952,796
 ; FILING DATE: 24-SEPT-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DOROTHY R. AUTH
 ; REGISTRATION NUMBER: 36434
 ; REFERENCE/DOCKET NUMBER: 202641260S1
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7210 Base Pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Double
 ; TOPOLOGY: Unknown
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Human
 ; IMMEDIATE SOURCE:
 ; LIBRARY: DASH II
 ; FEATURE:
 ; NAME/KEY: JTI06
 ; LOCATION:
 ; IDENTIFICATION METHOD:
 ; OTHER INFORMATION: 7.2 kb No. 5840686 1 fragments
 ; OTHER INFORMATION: Derived from human placental genomic DNA
 US-08-257-963B-10

Query Match 82.2%; Score 14.8; DB 2; Length 7210;
 Best Local Similarity 88.9%; Pred. No. 87;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccgaggtgggtggtgtaatt 18
 |||||
 DB 1617 CCGAGGTGGTGGATCAT 1634

RESULT 13
 PCT-US95-07201-10
 ; Sequence 10, Application PC/TUS9507201
 ; GENERAL INFORMATION:
 ; APPLICANT: Chader, Gerald J.; Becerra, Sofia
 ; APPLICANT: Patricia; Schwartz, Joan P.;
 ; APPLICANT: Taniwaki, Takayuki
 ; TITLE OF INVENTION: PIGMENT EPITHELIUM
 ; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
 ; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PRDF GENE
 ; NUMBER OF SEQUENCES: 43
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morgan & Finnegan, L.L.P.
 ; STREET: 345 Park Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/07201
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/367,841
 ; FILING DATE: 30-DEC-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/257,963
 ; FILING DATE: 07-JUN-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/952,796
 ; FILING DATE: 24-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DOROTHY R. AUTH
 ; REGISTRATION NUMBER: 36434
 ; REFERENCE/DOCKET NUMBER: 20264126PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7210 Base Pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Double
 ; TOPOLOGY: Unknown
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Human
 ; IMMEDIATE SOURCE:
 ; LIBRARY: DASH II
 ; FEATURE:
 ; NAME/KEY: JTI6A
 ; LOCATION:
 ; IDENTIFICATION METHOD:
 ; OTHER INFORMATION: 7.0 kb Not 1-Not
 ; OTHER INFORMATION: fragment; Derived from human placental
 ; OTHER INFORMATION: genomic DNA; also referred to as JTI06
 PCT-US95-07201-10

Query Match 82.2%; Score 14.8; DB 5; Length 7210;
 Best Local Similarity 88.9%; Pred. No. 87;

Thu Feb 21 07:26:09 2002

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
| | | | | | | | | | | | | | | | | |
Db 1617 CCGAGGTGGTGGTGCATC 1634

RESULT 14
PCT-US95-07201-43
; Sequence 43, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,841
; FILING DATE: 30-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126PCT
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Pl-147
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: full length genomic
; OTHER INFORMATION: sequence for PEDF plus flanking sequences.
PCT-US95-07201-43

Query Match 82.2%; Score 14.8; DB 5; Length 22481;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
| | | | | | | | | | | | | | | | | |
Db 1609 CCGAGGTGGTGGTGCATC 1626

RESULT 15
US-09-301-665-3
; Sequence 3, Application US/09301665
; Patent No. 6207876
; GENERAL INFORMATION:
; APPLICANT: KELLEMS, RODNEY E.
; APPLICANT: DATTA, SURJIT K.
; APPLICANT: BLACKBURN, MICHAEL R.
; TITLE OF INVENTION: ADENOSINE DEAMINASE DEFICIENT TRANSGENIC MICE AND
; TITLE OF INVENTION: METHODS FOR THE USE THEREOF
; FILE REFERENCE: UTSH:243
; CURRENT APPLICATION NUMBER: US/09/301,665
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: 60/083,408
; EARLIER FILING DATE: 1998-04-29
; EARLIER APPLICATION NUMBER: 60/083,370
; EARLIER FILING DATE: 1998-04-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 36741
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-301-665-3

Query Match 82.2%; Score 14.8; DB 4; Length 36741;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
| | | | | | | | | | | | | | | | | |
Db 2400 ccgaggtgggtgggtaat 2417

Search completed: February 20, 2002, 07:51:10
Job time: 578 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:25:49 ; Search time 3814.84 Seconds
(without alignments)
50.703 Million cell updates/sec

Title: US-09-904-420A-9
Perfect score: 18
Sequence: 1 ccgagggtgggtggaat 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_est1: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_vtr: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17	94.4	617	13	A0744329 HS_5502_A
c 2	16.4	91.1	181	10	AA573670 n142401-S
c 3	16.4	91.1	263	13	AQ217892 HS_2257-A
c 4	16.4	91.1	353	11	RG386035 602455204
c 5	16.4	91.1	359	13	R35371 HS-1028-B1-
6	16.4	91.1	382	11	T08252 EST06143 1u
7	16.4	91.1	399	11	BE940403 RC3-UT006
8	16.4	91.1	442	13	R33138 HS-1016-B1-
9	16.4	91.1	444	10	BE723699 193604 WA
c 10	16.4	91.1	455	13	B86654
11	16.4	91.1	470	13	B73071
12	16.4	91.1	471	13	AQ240635 CIT-HSP-2

c 13	16.4	91.1	471	13	AQ322652
c 14	16.4	91.1	492	13	AZ757574
c 15	16.4	91.1	508	13	AQ681066 HS_5492_B
c 16	16.4	91.1	512	13	AQ792852 HS_5253_B
c 17	16.4	91.1	593	13	AQ006721 CIT-HSP-2
c 18	16.4	91.1	831	11	BF130894 601819691
c 19	16.4	91.1	969	10	BE216883 HV_CEB001
c 20	16.4	91.1	970	11	BF682098 602117191
c 21	16	88.9	494	11	BF623736 HVSME000
c 22	16	88.9	797	11	BG344434 HVSME000
c 23	16	88.9	1144	10	BE213792 HV_CEB000
c 24	15.4	85.6	98	11	L76108 SCMRAP0193
c 25	15.4	85.6	216	11	BG950338 CM3-CT060
c 26	15.4	85.6	222	10	AI846112 UI-M-AK1-
c 27	15.4	85.6	233	10	AA320156 EST22475
c 28	15.4	85.6	240	10	BE164673 RC3-HT047
c 29	15.4	85.6	254	10	AA364999 EST75672
c 30	15.4	85.6	264	10	AA360997 EST70406
c 31	15.4	85.6	270	10	AA338386 EST43366
c 32	15.4	85.6	270	10	AA364466 EST75114
c 33	15.4	85.6	277	11	BF882491 CM1-ET019
c 34	15.4	85.6	282	10	BB148909 BB148909
c 35	15.4	85.6	288	10	BB307359 BB307359
c 36	15.4	85.6	290	10	BB004796 BB004796
c 37	15.4	85.6	291	10	AA350759 EST58292
c 38	15.4	85.6	292	10	AW841426 RC6-CN001
c 39	15.4	85.6	295	11	BG955500 CM4-CT065
c 40	15.4	85.6	304	10	BB262967 BB262967
c 41	15.4	85.6	306	10	BB500836 BB500836
c 42	15.4	85.6	307	10	BB269949 BB269949
c 43	15.4	85.6	316	10	AA564252 nK4912.S
c 44	15.4	85.6	317	11	BF853474 MR3-EN008
c 45	15.4	85.6	326	11	T59022 yb49D06.S1

ALIGNMENTS

RESULT 1
A0744329/c

LOCUS A0744329 617 bp DNA

DEFINITION HS_5502_A2_D04_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1078 Col=8 Row=G, DNA sequence.

ACCESSION A0744329

VERSION A0744329.1 GI:5521851

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 617)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GS, Wallace JC, Hood L.
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.husc.washington.edu
Plate: 1078 row: G column: 8
Seq primer: 17

Class: BAC ends

High quality sequence stop: 617.

Location/Qualifiers

FEATURES

source

1. .617

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="plate=1078 Col=8 Row=G"

/clone_lib="RPC1-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

146 a 130 c 122 g 216 t 3 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 94.4%; Score 17; DB 13; Length 617;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggttaa 17

|||||

Db 251 CCGAGGTGGGTGGGTAA 235

RESULT 2

AA573670/c

LOCUS

DEFINITION AA573670 181 bp mRNA EST 29-AUG-1997

n42901.sl NCI-CGAP_Pr2 Homo sapiens cDNA clone IMAGE:916464

similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA573670

VERSION AA573670.1

KEYWORDS GI:2348185

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 181)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.

, Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.

Clone distribution: Washington University Genome Sequencing Center

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/dbp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 162.

Location/Qualifiers

1. .181

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:916464"

/clone_lib="NCI-CGAP_Pr2"

/sex="Male"

/dev_stage="45 years old"

/lab_host="DH10B"

/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st

strand cDNA was primed with oligo(dT)17 on 50 ng of

DNase-treated, total cellular RNA obtained from 5,000-10

,000 microdissected preneoplastic cells

histologically-determined to be prostatic intraepithelial

neoplasia 2 (PIN2) cells. Double-stranded cDNA was

ligated to EcoRI adaptors. 5 cycles of PCR applied to the

cDNA with an adaptor-specific primer, and the resulting

PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman.

41 a 56 c 39 g 45 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 91.1%; Score 16.4; DB 10; Length 181;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggttaa 18

|||||

Db 152 CCGAGGTGGGTGGTCAT 135

RESULT 3

A0217892

LOCUS

DEFINITION A0217892 263 bp DNA GSS 19-SEP-1998

HS_2257_B2_COL_T7 CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=2257 Col=2 Row=F, DNA sequence.

ACCESSION A0217892

VERSION A0217892.1

KEYWORDS GI:3629519

SOURCE GSS.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 263)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,L.S., Holzman,T., and

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,W.D., and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2257 row: F column: 2

Class: BAC ends

High quality sequence stop: 263.

Location/Qualifiers

1. .263

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=2257 Col=2 Row=F"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="Male"

/note="Organ: sperm; Vector: pBIOBAC11; HAC Clones in

E-Coli DH10B"

62 a 61 c 64 g 74 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 91.1%; Score 16.4; DB 13; Length 263;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggttaa 18

|||||

Db 151 CCGAGGTGGGTGGTCAT 168

RESULT 4

BG386035/c

LOCUS

BG386035

353 bp mRNA

EST

12-MAR-2001

```

DEFINITION 602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5',
mRNA sequence.
ACCESSION BG386035
VERSION BG386035.1 GI:13279481
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 353)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1307 row: h column: 02
High quality sequence stop: 276.
FEATURES             source
   source
   1..353
   Location/Qualifiers
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /clone="IMAGE:4583473"
       /clone_lib="NIH_MGC_15"
       /tissue_type="adenocarcinoma cell line"
       /lab_host="DH10B (phage-resistant)"
       /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
       EcoRI; cDNA made by oligo-dT priming. Directionally
       cloned into EcoRI/XhoI sites using the following 5',
       adaptor: GGCACGAG(C). Size-selected >500bp for average
       insert size 1.8kb. Library constructed by Ling Hong in
       the laboratory of Gerald M. Rubin (University of
       California, Berkeley) using ZAP-cDNA synthesis kit
       (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      80 a 111 c 122 g 40 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 11; Length 353;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
||||| |||||||||
Db 262 CCGAGGGGGGTGGGTAAT 245

RESULT 5
LOCUS B35371/c 359 bp DNA GSS 17-OCT-1997
DEFINITION HS-1028-B1-E06-MR.ab1 CIT Human Genomic Sperm Library C Homo
sapiens genomic clone Plate-CT 807 Col-11 Row-J, DNA sequence.
ACCESSION B35371
VERSION B35371.1 GI:2534740
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
Traiicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.F.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L.

```

```

University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackron@u.washington.edu
Sequence Tagged Connector
Plate: CT 807, row: J column: 11
Class: RAC ends
High quality sequence stop: 359.
Location/Qualifiers
   1..359
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="Plate-CT 807 Col-11 Row-J"
   /clone_lib="CIT Human Genomic Sperm Library C"
   /sex="M"
   /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
   E-Coli DH10B"
BASE COUNT      72 a 101 c 84 g 102 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 13; Length 359;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
||||| |||||||||
Db 303 CCGAGGTGGGTGGTCAT 286

RESULT 6
LOCUS T08252 382 bp mRNA EST 03-AUG-1993
DEFINITION EST06143 Infant Brain, Bento Soares Homo sapiens cDNA clone H1BBB329
5' end similar to EST containing Alu repeat, mRNA sequence.
ACCESSION T08252
VERSION T08252.1 GI:389280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 382)
AUTHORS Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018669056
Fax: 3018669423
Email: mdamad@etigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
   1..382
   /organism="Homo sapiens"
   /db_xref="ATCC (inhost):84798"
   /db_xref="taxon:9606"
   /clone="H1BBB329"
BASE COUNT      98 a 81 c 110 g 90 t 3 others
ORIGIN
Query Match      91.1%; Score 16.4; DB 11; Length 382;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccgaggtgggtgggtaat 18
||||| |||||||||

```



```

Db 158 CCGAGTGGTGGGTAT 175

RESULT 7
BE940403
LOCUS RC3-UT0062-210800-011-e01 UT0062 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE940403 EST 02-OCT-2000
ACCESSION BE940403.1 GI:10469865
VERSION BE940403.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 399)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL.
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-RC3-UT0062-210
800-011-e01&t3=2000-08-21&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 399.
FEATURES
source
Location/Qualifiers
1..399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0062"
/dev_stage="Adult"
/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 102 a 79 c 98 g 120 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 11; Length 399;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggtggaat 18
|||||
Db 183 CCGAGTGGTGGGTAT 200

RESULT 8
B33138
LOCUS HS-1016-B1-E02-MR.abi CIT Human Genomic Sperm Library C Homo
DEFINITION B33138 DNA 17-OCT-1997
ACCESSION B33138 GSS
VERSION B33138
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 442)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL.
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-RC3-UT0062-210
800-011-e01&t3=2000-08-21&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 399.
FEATURES
source
Location/Qualifiers
1..442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelorBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 95 a 92 c 105 g 150 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 442;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggtggaat 18
|||||
Db 397 CTGAGTGGTGGGTAT 414

RESULT 9
BE723699
LOCUS I93604 MARC 280V Bos Laurus cDNA 5', mRNA sequence.
DEFINITION BE723699 EST 25-APR-2001
ACCESSION BE723699.1 GI:10124995
VERSION BE723699.1
KEYWORDS EST.
SOURCE COM.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bos.
1 (bases 1 to 444)
Smith,T.P.L., Grosse,W.M., Preking,B.A., Roberls,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Hennictt
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perte,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPI.
USDA, ARS, US Meat Animal Research Center,
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

```

```

v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
BACKWARD: AGGAACACCTATGACCAT
Plate: 38 row: I column: 18
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .444
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2B0Y"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendonosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT      80 a 147 c 134 g 83 t
ORIGIN

Query Match      91.1%; Score 16.4; DB 10; Length 444;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggtaaat 18
|||||
Db 171 CCGAGTGGGTGGGTGAT 188

RESULT 10
B86654/c
LOCUS      B86654      455 bp      DNA      GSS      09-APR-1999
DEFINITION RPC111-24H22.TP RPC1-11 Homo sapiens genomic clone RPC1-11-24H22,
DNA sequence.
ACCESSION  B86654
VERSION    B86654.1 GI:2927786
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 455)
AUTHORS    Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
            Use of BAC End Sequences for Sequence-Ready Map Building (1998)
            Unpublished (1998)
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPC1-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
            Location/Qualifiers
            source
            1. .455
            /organism="Homo sapiens"
            /db_xref="GDB:7509021"
            /db_xref="taxon:9606"
            /clone_lib="RPC1-11-24H22"
            /sex="Male"
            /cell_type="Lymphocytes"

FEATURES
            source
            1. .455
            /organism="Homo sapiens"
            /db_xref="GDB:7509021"
            /db_xref="taxon:9606"
            /clone_lib="RPC1-11-24H22"
            /sex="Male"
            /cell_type="Lymphocytes"

/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT      110 a 119 c 86 g 140 t
ORIGIN

Query Match      91.1%; Score 16.4; DB 13; Length 455;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggtaaat 18
|||||
Db 435 CCGAGTGGGTGGGTGAT 418

RESULT 11
B73071
LOCUS      B73071      470 bp      DNA      GSS      08-APR-1999
DEFINITION RPC111-11C13.TP RPC1-11 Homo sapiens genomic clone RPC1-11-11C13,
DNA sequence.
ACCESSION  B73071
VERSION    B73071.1 GI:2712222
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 470)
AUTHORS    Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
            Use of BAC End Sequences for Sequence-Ready Map Building
            Unpublished (1997)
            Other_GSSs: RPC111-11C13.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPC1-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
            Location/Qualifiers
            source
            1. .470
            /organism="Homo sapiens"
            /db_xref="GDB:7503900"
            /db_xref="taxon:9606"
            /clone_lib="RPC1-11-11C13"
            /clone_lib="RPC1-11"
            /sex="Male"
            /cell_type="Lymphocytes"
            /notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPC111 Human Male BAC Library"
BASE COUNT      158 a 98 c 102 g 112 t
ORIGIN

Query Match      91.1%; Score 16.4; DB 13; Length 470;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccgagtggtgggtaaat 18
|||||
Db 52 CCGAGTGGGTGGGTGAT 69

```

```

RESULT 12
AQ240635      471 bp      DNA      GSS      30-SEP-1998
LOCUS
DEFINITION    CIT-HSP-2385B22.TF.1 CIT-HSP Homo sapiens genomic clone 2385B22,
DNA sequence.
ACCESSION     AQ240635
VERSION       AQ240635
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 471)
AUTHORS       Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE         Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL
COMMENT       Other_GSSs: CIT-HSP-2385B22.TR.1
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M3-21
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..471
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="CIT-HSP"
                /sex="Male"
                /cell_type="Sperm"
                /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                HindIII"
BASE COUNT    158 a    90 c    126 g    97 t
ORIGIN
    Query Match      91.1%; Score 16.4; DB 13; Length 471;
    Best Local Similarity 94.4%; Pred. No. 2.1e+03;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ccgaggtgggtggtgtaat 18
    ||| ||||| ||||| |||||
Db 118 CCAAGTGGGTGGTAAAT 135

RESULT 13
AQ322652/c    471 bp      DNA      GSS      06-MAY-1999
LOCUS
DEFINITION    RPC111-107H7.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-107H7,
DNA sequence.
ACCESSION     AQ322652
VERSION       AQ322652
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 471)
AUTHORS       Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE         Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL
COMMENT       Other_GSSs: RPC111-107H7.TV

```

```

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pjeter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES
    Location/Qualifiers
        1..471
            /organism="Homo sapiens"
            /db_xref="GDB:7540878"
            /db_xref="taxon:9606"
            /clone="RPC1-11-107H7"
            /clone_lib="RPC1-11"
            /sex="Male"
            /cell_type="Lymphocytes"
            /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPC111 Human Male BAC Library"
BASE COUNT    90 a    120 c    101 g    160 t
ORIGIN
    Query Match      91.1%; Score 16.4; DB 13; Length 471;
    Best Local Similarity 94.4%; Pred. No. 2.1e+03;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ccgaggtgggtggtgtaat 18
    ||| ||||| ||||| |||||
Db 322 CCGAGTGGGTGGTAAAT 305

RESULT 14
AZ757574/c    492 bp      DNA      GSS      01-MAR-2001
LOCUS
DEFINITION    ew09h05.r1 PAX3/EKHR CASTing Library 'ew' Homo sapiens genomic
clone ew09h05 random, DNA sequence.
ACCESSION     AZ757574
VERSION       AZ757574.1 GI:13177040
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 492)
AUTHORS       Barber,T.D., Barber,M.C., Tomescu,O., Barr,F., Ruben,S. and
Friedman,T.B.
TITLE         Cyclic amplification and selection of target genes regulated by
Pax3 and PAX3/EKHR in embryogenesis and alveolar rhabdomyosarcoma
JOURNAL
COMMENT       Unpublished (2000)
Contact: Friedman TB
Laboratory of Molecular Genetics
National Institute on Deafness and Other Communication Disorders,
National Institutes of Health
5 Research Court, Room 2A-15, Rockville, MD 20850, USA
Tel: 301 402 7580
Fax: 301 496 7882
Email: friedman@nidcd.nih.gov
Plate: 09 row: h column: 05
Seq primer: -28M13 reverse primer (Amersham)
Class: random plasmid subclone.
FEATURES
    Location/Qualifiers
        1..492
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="ew09h05"

```

/clone_lib="PAX3/FKHR CASTING Library 'ew'"
/sex="Male"

/lab_host="DH10B"
/note="Vector: pGEM-T Easy: Human genomic DNA was
partially digested with Sau3AI, ligated to ds linkers,
and enriched for binding to human PAX3/FKHR protein using
a whole genome PCR-based strategy. DNA fragments
containing putative PAX3/FKHR binding sites were
amplified by PCR and cloned into pGEM-T Easy (Promega).
The ligation products were transformed into DH10B
electrocompetent cells (Life Technologies)."

BASE COUNT 118 a 129 c 102 g 143 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13: Length 492;
Best Local Similarity 94.4%; Pred. NO. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ccgaggtgggtgggtaat 18
|||||
Db 320 CCGAGGTGGGTGGGTAAT 303

RESULT 15
AQ681066/c
LOCUS
DEFINITION HS_5492.B2_H03.T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1068 Col-6 Row=P, DNA sequence.
ACCESSION AQ681066
VERSION
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 508)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://www.htsc.washington.edu>)
<http://www.htsc.washington.edu>
Plate: 1068 row: P column: 6
Seq primer: T7
Class: BAC ends
High quality sequence stop: 508.
Location/Qualifiers

FEATURES
source
1..508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1068 Col-6 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 100 a 141 c 104 g 159 t 4 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 13: Length 508;
Best Local Similarity 94.4%; Pred. NO. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ccgaggtgggtgggtaat 18
|||||
Db 113 CCGAGGTGGGTGGGTCAT 96

Search completed: February 20, 2002, 10:25:52
Job time: 9860 sec

us-09-904-420a-9.rst

Thu Feb 21 07:26:11 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 11:01:45 ; Search time 1729.17 Seconds
(without alignments)
190.810 Million cell updates/sec

Title: US-09-904-420A-8

Perfect score: 20

Sequence: 1 attaccaccccccctcgga 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_ov:*
- 21: em_pat:*
- 22: em_ph:*
- 23: em_pi:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	20	100.0	1115	9	HSTCFID	Z47361 H.sapiens T
2	20	100.0	1165	9	HSTCFIE	Z47362 H.sapiens T
3	20	100.0	1254	9	HSTCFIA	X59869 Human TCF-1
4	20	100.0	2814	9	HSTCFIB	X59870 Human TCF-1
5	20	100.0	2855	9	HSTCFIG	X63901 Homo sapien
6	20	100.0	2910	9	HSTCFIC	X59871 Human TCF-1
c 7	20	100.0	84544	2	AC009012	AC009012 Homo sapi
c 8	20	100.0	86914	2	AC011336	AC011336 Homo sapi
c 9	20	100.0	159420	2	AC009017	AC009017 Homo sapi
c 10	20	100.0	200831	2	AC008608	AC008608 Homo sapi
c 11	17.4	87.0	112209	10	AC022298	AC022298 Mus muscu
c 12	17.4	87.0	214780	2	AC069075	AC069075 Mus muscu
c 13	17	85.0	136328	9	AC004126	AC004126 Human Chr
c 14	17	85.0	158991	2	AC087673	AC087673 Homo sapi
c 15	17	85.0	161245	2	AP001986	AP001986 Homo sapi
c 16	17	85.0	195616	2	AC090344	AC090344 Homo sapi
c 17	17	85.0	197156	2	AC090345	AC090345 Homo sapi
c 18	17	85.0	198241	2	AP003721	AP003721 Homo sapi
c 19	16.8	84.0	66256	2	AC087454	AC087454 Homo sapi
c 20	16.8	84.0	148699	2	AC092736	AC092736 Trypanoso
c 21	16.8	84.0	153241	2	AC024960	AC024960 Homo sapi
c 22	16.8	84.0	170537	2	AC073548	AC073548 Homo sapi
c 23	16.4	82.0	631	9	AB046035	AB046035 Macaca fa
c 24	16.4	82.0	53721	2	AC013963	AC013963 Drosophil
c 25	16.4	82.0	138634	2	AP003846	AP003846 Oryza sat
c 26	16.4	82.0	154115	9	AC010878	AC010878 Homo sapi
c 27	16.4	82.0	157939	2	AL359384	AL359384 Homo sapi
c 28	16.4	82.0	159855	2	AC079883	AC079883 Rattus no
c 29	16.4	82.0	167912	3	AC007650	AC007650 Drosophil
c 30	16.4	82.0	173738	2	AC069293	AC069293 Homo sapi
c 31	16.4	82.0	173965	9	AC018890	AC018890 Homo sapi
c 32	16.4	82.0	178405	2	AC025927	AC025927 Homo sapi
c 33	16.4	82.0	179040	2	AC022732	AC022732 Homo sapi
c 34	16.4	82.0	186827	2	AC026478	AC026478 Mus muscu
c 35	16.4	82.0	188540	2	AC011108	AC011108 Homo sapi
c 36	16.4	82.0	193991	2	AC090432	AC090432 Mus muscu
c 37	16.4	82.0	202915	2	AC060773	AC060773 Homo sapi
c 38	16.4	82.0	225974	3	AE003702	AE003702 Drosophil
c 39	16	80.0	127913	9	AL590727	AL590727 Human DNA
c 40	16	80.0	132835	2	AC010059	AC010059 Drosophil
c 41	16	80.0	159057	2	AC023995	AC023995 Homo sapi
c 42	16	80.0	168258	2	AC018703	AC018703 Homo sapi
c 43	16	80.0	190489	9	AC067794	AC067794 Homo sapi
c 44	16	80.0	205430	2	AP002756	AP002756 Homo sapi
c 45	16	80.0	211325	2	AP002757	AP002757 Homo sapi

ALIGNMENTS

RESULT 1

HSTCFID

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

HSTCFID 1115 bp mRNA
H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.
247361
247361.1 GI:619881
splice form D; T cell factor 1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1115)
Mayer K., Wolff E., Clevers H. and Ballhausen W.G.
The human high mobility group (HMG)-box transcription factor TCF-1:
novel isoforms due to alternative splicing and usage of a new exon
IXA
Unpublished
2 (bases 1 to 1115)
van de Wetering M., Oosterwegel M., Holsteg F., Dooyes D.,
Suijkerbuijk K., Geurts van Kessel A. and Clevers H.

```

TITLE      The human T cell transcription factor-1 gene. Structure,
JOURNAL    localization, and promoter characterization
MEDLINE    J. Biol. Chem. 267 (12), 8530-8536 (1992)
REFERENCE  92235082
AUTHORS    3 (bases 1 to 1115)
TITLE      Ballhausen, W.G.
JOURNAL    Direct Submission
            Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
            Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
            Schwabachanlage 10, Erlangen, Germany, D-91054
FEATURES   Location/Qualifiers
            1..1115
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="lymphoma"
            /cell_type="T-lymphocyte"
            /cell_line="Jurkat"
            1..1115
            /gene="TCF-1"
            /standard_name="T cell factor 1 splice form D"
            /citation=[1]
            /function="high mobility group box transcription factor"
            /evidence=experimental
            1..1115
            /gene="TCF-1"
            2..754
            /standard_name="T cell factor 1 splice form D"
            /citation=[1]
            /codon_start=1
            /product="T cell factor 1 splice form D"
            /protein_id="CAA87439.1"
            /db_xref="GI:619882"
            /translation="MYKTVYSAFNLLMHYPSPGAGQHPPQPLHKAQPPHGVPO
            LSLVEFNSPHPTPADISOKVHRPLOTDLGFSYSLTSGMGQLPHTVSWFTHPS
            LMLGSGVGPHPAAIPHPAIVPPSGKQELQPDRLNKTQAESKAKEAKKPTIKKPLNA
            FMYLMKEMRAKVIABCTLKESAAINQILGRWHALSREEQAKYELARKERQLHMQLY
            PGWSARDNTYKKRRSREKHQESTTDSLSHYS"
            731..751
            mat_peptide
            /gene="TCF-1"
            /product="alternative ORF specific for TCF-1D"
            272 a 407 c 265 g 171 t

BASE COUNT 272 a 407 c 265 g 171 t

Query Match 100.0%; Score 20; DB 9; Length 1115;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attaccacccccctcgga 20
Db 45 ATTACCACCCCCCTCGGA 64

RESULT 2
HSTCF1E HSTCF1E 1165 bp mRNA PRI 09-JAN-1995
LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.
DEFINITION
ACCESSION Z47362
VERSION Z47362.1 GI:619883
KEYWORDS splice form E; T cell factor 1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.
TITLE The human high mobility group (HMG)-box transcription factor TCF-1:
        novel isoforms due to alternative splicing and usage of a new exon
        1A
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1165)
AUTHORS van de Wetering, M., Oosterwegel, M., Holstege, F., Duoyes, D.,

```

```

TITLE      The human T cell transcription factor-1 gene. Structure,
JOURNAL    localization, and promoter characterization
MEDLINE    J. Biol. Chem. 267 (12), 8530-8536 (1992)
REFERENCE  92235082
AUTHORS    3 (bases 1 to 1165)
TITLE      Ballhausen, W.G.
JOURNAL    Direct Submission
            Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
            Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
            Schwabachanlage 10, Erlangen, Germany, D-91054
FEATURES   Location/Qualifiers
            1..1165
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="lymphoma"
            /cell_type="T-lymphocyte"
            /cell_line="Jurkat"
            1..1165
            /gene="TCF-1"
            /standard_name="T cell factor 1 splice form E"
            /citation=[1]
            /function="high mobility group box transcription factor"
            /evidence=experimental
            1..1165
            /gene="TCF-1"
            2..1144
            /standard_name="T cell factor 1 splice form E"
            /citation=[1]
            /codon_start=1
            /product="T cell factor 1 splice form E"
            /protein_id="CAA87440.1"
            /db_xref="GI:619884"
            /translation="MYKTVYSAFNLLMHYPSPGAGQHPPQPLHKAQPPHGVPO
            LSLVEFNSPHPTPADISOKVHRPLOTDLGFSYSLTSGMGQLPHTVSWFTHPS
            LMLGSGVGPHPAAIPHPAIVPPSGKQELQPDRLNKTQAESKAKEAKKPTIKKPLNA
            FMYLMKEMRAKVIABCTLKESAAINQILGRWHALSREEQAKYELARKERQLHMQLY
            PGWSARDNTYKKRRSREKHQESTTDPGSKKCRARFLNQOTDWCPCRRKKCIKY
            LPGEKCPSPVPSDDSGALGCPGAPQSPSYHLLPRPTFILTSPAERHLHPQVSHL
            LSASOPGPHRPPAAPACRAHRYSNRNLDRWPSRHRTPGRLOEPTP"
            803..1141
            mat_peptide
            /gene="TCF-1"
            /product="alternative ORF specific for TCF-1E"
            279 a 422 c 286 g 178 t

BASE COUNT 279 a 422 c 286 g 178 t

Query Match 100.0%; Score 20; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attaccacccccctcgga 20
Db 45 ATTACCACCCCCCTCGGA 64

RESULT 3
HSTCF1A HSTCF1A 1254 bp mRNA PRI 14-JUN-1991
LOCUS human TCF-1 mRNA for T cell factor 1 (splice form A).
DEFINITION
ACCESSION X59869
VERSION X59869.1 GI:36785
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van de Wetering, Dept of Clinical
        Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

```

REFERENCE 2 (bases 1 to 1254)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooljes, D. and Clevers, H.
TITLE Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box
JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 91114695
COMMENT See also X59869-X59871.
FEATURES Location/Qualifiers
 source 1..1254
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="T lymphocyte"
 /cell_line="Jurkat"
 /clone_lib="cDNA"
 /clone_lib="fTCF-1a"
 mRNA 1..1254
 /gene="TCF-1"
 /evidence="experimental"
 gene 1..1254
 /gene="TCF-1"
 CDS 80..889
 /codon_start=1
 /product="T cell factor 1, splice form A"
 /protein_id="CAA42526.1"
 /db_xref="GI:36786"
 /db_xref="SWISS-PROT:P36402"
 /translation="MYKTVYSAPNLLMIYPPSGAGHPQPOPLHKANOPPHGVPO
LSLYEHFNSPHTPAADISOKVHRPLQTDLSGFYSLTSGMGLPHVSWFTHPS
LMIGSGVPGHPAAIPHPAIVPPSGKQELQPPDRNI.KTOAESKAKEAKKPTIKKPLNA
FMLYMKEMRAKVIAECTLKESAAINQILGRWHALSREOAKYVELAKKERQLHMOLY
PGWSARDNTGKKRRSRKREKHQESTTETNPRELKDGNGQESLSMSSSSPA"
 /misc_feature 539..769
 /gene="TCF-1"
 /note="HMG box"
 misc_feature 810
 /gene="TCF-1"
 /note="alternative splice site"
BASE COUNT 320 a 391 c 335 g 208 t
ORIGIN
Query Match 100.0%; Score 20; DB 9; Length 1254;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attaccacccccctcgga 20
|||||
DB 123 ATTACCCACCCCTCGGA 142
RESULT 4
HSTCF1B HSTCF1B 2814 bp mRNA PRI 17-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form H).
DEFINITION X59870 X55329
ACCESSION X59870.1 GI:36787
VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2814)
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht, The Netherlands
REFERENCE 2 (bases 1 to 2814)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooljes, D. and Clevers, H.
TITLE Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box

JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 91114695
COMMENT See also X59869-X59871.
FEATURES Location/Qualifiers
 source 1..2814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="T lymphocyte"
 /cell_line="Jurkat and H9B-ALL"
 /clone_lib="cDNA"
 mRNA 1..2814
 /gene="TCF-1"
 /evidence="experimental"
 gene 1..2814
 /gene="TCF-1"
 CDS 80..889
 /gene="TCF-1"
 /codon_start=1
 /product="T cell factor 1 splice form B"
 /protein_id="CAA42527.1"
 /db_xref="GI:36788"
 /db_xref="SWISS-PROT:P36402"
 /translation="MYKTVYSAPNLLMIYPPSGAGHPQPOPLHKANOPPHGVPO
LSLYEHFNSPHTPAADISOKVHRPLQTDLSGFYSLTSGMGLPHVSWFTHPS
LMIGSGVPGHPAAIPHPAIVPPSGKQELQPPDRNI.KTOAESKAKEAKKPTIKKPLNA
FMLYMKEMRAKVIAECTLKESAAINQILGRWHALSREOAKYVELAKKERQLHMOLY
PGWSARDNTGKKRRSRKREKHQESTTGGKKNAGCTYPEKANAAPFLPWTVL"
 /misc_feature 539..769
 /gene="TCF-1"
 /note="HMG box"
 misc_feature 810
 /gene="TCF-1"
 /note="alternative splice site"
BASE COUNT 763 a 831 c 648 g 572 t
ORIGIN
Query Match 100.0%; Score 20; DB 9; Length 2814;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attaccacccccctcgga 20
|||||
DB 123 ATTACCCACCCCTCGGA 142
RESULT 5
HSTCF1G HSTCF1G 2855 bp DNA PRI 30-SEP-1999
LOCUS Homo sapiens TCF-1 gene.
DEFINITION X63901
ACCESSION X63901.1 GI:36791
VERSION T-cell transcription factor; transcription factor.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2855)
AUTHORS van de Wetering, M., Oosterwegel, M., Holtsiege, F., Dooyes, D.,
Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.
TITLE The human T cell transcription factor-1 gene. Structure,
localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
JOURNAL 92235082
MEDLINE 2 (bases 1 to 2855)
AUTHORS van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H.
TITLE Extensive alternative splicing and dual promoter usage generate
Tcf-1 protein isoforms with differential transcription control
properties
Mol. Cell. Biol. 16 (3), 745-752 (1996)
JOURNAL 96182076
MEDLINE 3 (bases 1 to 2855)
AUTHORS van de Wetering, M.L.

RESULT 9
AC009017/C

LOCUS
DEFINITION

AC009017 159420 bp DNA HTG 20-APR-2001
Homo sapiens chromosome 5 clone XXp1-929G6, WORKING DRAFT SEQUENCE,
26 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE

AC009017 GI:13699469
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 159420)

REFERENCE
AUTHORS
TITLE

DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished

REFERENCE
AUTHORS
TITLE

DOE Joint Genome Institute.
Direct Submission

JOURNAL

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:1178048.

COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 1189133, H51
Center clone name: XXp1-929G6

Summary Statistics

Consensus quality: 142845 bases at least Q40
Consensus quality: 150910 bases at least Q30
Consensus quality: 152756 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 156920; sum-of-contigs estimation
Quality coverage: 8.41 in Q20 bases; pulse field gel estimation.
Quality coverage: 4.29 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1423: contig of 1423 bp in length
1424 1523: gap of unknown length
1524 3066: contig of 1543 bp in length
3067 3166: gap of unknown length
3167 5623: contig of 2463 bp in length
5630 5723: gap of unknown length
5730 6997: contig of 1268 bp in length
6998 7097: gap of unknown length
7098 9214: contig of 2117 bp in length
9215 9315: gap of unknown length
9316 11813: contig of 2499 bp in length
11814 11913: gap of unknown length
11914 14718: contig of 2805 bp in length
14719 14813: gap of unknown length
14819 17174: contig of 2356 bp in length
17175 17274: gap of unknown length
17275 19710: contig of 2436 bp in length
19711 19810: gap of unknown length
19811 21678: contig of 1868 bp in length
21679 21778: gap of unknown length
21779 26021: contig of 4243 bp in length
26022 26121: gap of unknown length
26122 29609: contig of 3488 bp in length
29610 33196: gap of unknown length
33197 33296: gap of unknown length

FEATURES

source

1..159420
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXp1-929G6"

BASE COUNT 39367 a 38414 c 39218 g 39915 t 2506 others
ORIGIN

Query Match 100.0%; Score 20; DB 2; Length 159420;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attaccacccccctcgga 20

|||||

Db 56825 ATTACCACCCCTCGGA 56806

RESULT 10

AC008608

LOCUS

DEFINITION

AC008608 200831 bp DNA HTG 20-APR-2001

Homo sapiens chromosome 5 clone CTB-113I20, WORKING DRAFT SEQUENCE,
29 unordered pieces.

ACCESSION

AC008608

VERSION

AC008608.5 GI:13699394

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 200831)

AUTHORS

DOE Joint Genome Institute.

TITLE

Sequencing of Human Chromosome 5

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 200831)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7708985.

COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 107808, H233

Center clone name: CIT978SKB_113120

Summary Statistics
Consensus quality: 174733 bases at least Q40
Consensus quality: 186693 bases at least Q30
Consensus quality: 190547 bases at least Q20
Estimated insert size: 165000; pulse field gel estimation
Quality coverage: 198031; sum-of-contigs estimation
Quality coverage: 7.77 in Q20 bases; pulse field gel estimation
Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1115: contig of 1115 bp in length
* 1116: 1215: gap of unknown length
* 1216: 2567: contig of 1352 bp in length
* 2568: 2667: gap of unknown length
* 2668: 3745: contig of 1078 bp in length
* 3746: 3845: gap of unknown length
* 3846: 4920: contig of 1075 bp in length
* 4921: 5020: gap of unknown length
* 5021: 6044: contig of 1024 bp in length
* 6045: 6144: gap of unknown length
* 6145: 7285: contig of 1141 bp in length
* 7286: 7385: gap of unknown length
* 7386: 8659: contig of 1274 bp in length
* 8660: 8759: gap of unknown length
* 8760: 9804: contig of 1045 bp in length
* 9805: 11696: contig of 1792 bp in length
* 11697: 11796: gap of unknown length
* 11797: 12841: contig of 1045 bp in length
* 12842: 12941: gap of unknown length
* 12942: 15043: contig of 2102 bp in length
* 15044: 15143: gap of unknown length
* 15144: 17289: contig of 2146 bp in length
* 17290: 17389: gap of unknown length
* 17390: 20232: contig of 2843 bp in length
* 20233: 20332: gap of unknown length
* 20333: 21884: contig of 1552 bp in length
* 21885: 21984: gap of unknown length
* 21985: 24278: contig of 2294 bp in length
* 24279: 24378: gap of unknown length
* 24379: 28103: contig of 3725 bp in length
* 28104: 28203: gap of unknown length
* 28204: 30867: contig of 2664 bp in length
* 30868: 30967: gap of unknown length
* 30968: 34914: contig of 3947 bp in length
* 34915: 35014: gap of unknown length
* 35015: 36030: contig of 1016 bp in length
* 36031: 36130: gap of unknown length
* 36131: 38852: contig of 2722 bp in length
* 38853: 38952: gap of unknown length
* 38953: 41871: contig of 2919 bp in length
* 41872: 41971: gap of unknown length
* 41972: 47080: contig of 5109 bp in length
* 47081: 47180: gap of unknown length
* 47181: 49750: contig of 2570 bp in length
* 49751: 49850: gap of unknown length
* 49851: 57118: contig of 7268 bp in length
* 57119: 57218: gap of unknown length
* 57219: 66046: contig of 8828 bp in length
* 66047: 66146: gap of unknown length
* 66147: 81046: contig of 14900 bp in length
* 81047: 81146: gap of unknown length
* 81147: 101650: contig of 20504 bp in length
* 101651: 101750: gap of unknown length
* 101751: 137108: contig of 35358 bp in length

* 137109 137208: gap of unknown length
* 137209 200831: contig of 63623 bp in length.
FEATURES
Source
Location/Qualifiers
1. .200831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone=CFB-113120"
/clone_lib="Caltech human BAC library B"
BASE COUNT 53266 a 46907 c 47140 g 50695 t 2823 others
ORIGIN
Query Match 100.0%; Score 20; PA 2; Length 200831;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attaccacccccctcgga 20
|||||
Db 184029 ATTACCACCCCTCGGA 184048
RESULT 11
AC022298
LOCUS BAC Library complete sequence.
DEFINITION Mus musculus BAC RP23-423C8 (Roswell Park Cancer Institute Mouse
AC022298
ACCESSION AC022298.31 GI:12621354
KEYWORDS HTG
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 112209)
AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Ransey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogue, M.,
Hosak, H., Hou, J., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louised, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, C., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.
Direct Submission
TITLE Unpublished
JOURNAL 2 (bases 1 to 112209)
REFERENCE 2 (bases 1 to 112209)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 112209)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 112209)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 112209)
AUTHORS Worley, K.C.
TITLE Direct Submission

JOURNAL Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jan 31, 2001 this sequence version replaced gi:12408434. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

source 1..112209
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-423C8"
 complement(1..215)
 /note="Overlaps bases 215..1 of clone AC074046"
 complement(833..904)
 /rpt_family="L1_MM"
 2174..2251
 /rpt_family="(TA)n"
 2639..2956
 /rpt_family="ORR1A2"
 4298..4355
 /rpt_family="(CCA)n"
 6047..6130
 /rpt_family="(TG)n"
 complement(6131..6323)
 /rpt_family="Lx8"
 6806..7245
 /rpt_family="Lx2"
 complement(9286..9580)
 /rpt_family="L2"
 9939..9966
 /rpt_family="AT_rich"
 10816..11100
 /note="Region similar to Mm#S793070 AV271329 Mus musculus
 cDNA: AV271329"
 10849..11098
 /note="Region similar to Mm#S793005 AV271264 Mus musculus
 cDNA: AV271264"
 complement(11142..11394)
 /rpt_family="Lx2"
 complement(11428..11501)
 /rpt_family="Lx4"
 11500..11830
 /rpt_family="Lx2"
 12011..12175
 /rpt_family="RSINEL"
 12590..12714
 /rpt_family="(TAGA)n"
 12725..12763
 /rpt_family="(CA)n"
 13082..13358
 /rpt_family="L2"
 14419..14451
 /rpt_family="AT_rich"
 14684..14750
 /rpt_family="(ATTG)n"
 15979..16008
 /rpt_family="AT_rich"
 16114..16148
 /rpt_family="AT_rich"

repeat_region complement(16199..16311)
 /rpt_family="B2"
 16540..16562
 /rpt_family="AT_rich"
 17041..17090
 /rpt_family="(TA)n"
 complement(17212..17279)
 /rpt_family="Lx2"
 17893..17918
 /rpt_family="(TTG)n"
 complement(17930..18046)
 /rpt_family="Lx"
 complement(18243..18310)
 /rpt_family="B2"
 complement(18366..18540)
 /rpt_family="B2"
 complement(18620..18722)
 /rpt_family="MIR"
 19102..19165
 /rpt_family="(CA)n"
 19933..20265
 /rpt_family="Lx2"
 21405..21426
 /rpt_family="AT_rich"
 21787..21809
 /rpt_family="(CAAA)n"
 complement(25271..25515)
 /rpt_family="L1_MM"
 25525..25793
 /rpt_family="L1_MM"
 25815..25844
 /rpt_family="(CAAAA)n"
 26118..26185
 /rpt_family="T-rich"
 26399..26696
 /rpt_family="L1_MM"
 26910..26986
 /rpt_family="(GA)n"
 27018..27070
 /rpt_family="GA-rich"
 27440..27501
 /rpt_family="(CA)n"
 27900..27922
 /rpt_family="(T)n"
 28340..28431
 /rpt_family="L2"
 28503..28587
 /rpt_family="PB1D10"
 28639..28701
 /rpt_family="(TAGA)n"
 28808..29989
 /rpt_family="Lx2"
 complement(30545..30648)
 /rpt_family="MIR"
 33954..33978
 /rpt_family="(T)n"
 34115..34151
 /rpt_family="AT_rich"
 34837..34882
 /rpt_family="(TTTA)n"
 complement(34932..35077)
 /rpt_family="B3"
 35102..35128
 /rpt_family="AT_rich"
 complement(35995..36178)
 /rpt_family="B3A"
 36181..36237
 /rpt_family="(TTTTG)n"
 36254..36317
 /rpt_family="T-rich"
 complement(36320..36587)
 /rpt_family="MTC"
 36621..36671
 repeat_region

```

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 41075: contig of 41075 bp in length
* 41076: gap of unknown length
* 41176: contig of 39764 bp in length
* 80939: contig of unknown length
* 80940: gap of unknown length
* 81040: contig of 27705 bp in length
* 108744: gap of unknown length
* 108745: contig of 28869 bp in length
* 108845: gap of unknown length
* 137113: contig of 28869 bp in length
* 137114: gap of unknown length
* 137813: contig of 28869 bp in length
* 137814: gap of unknown length
* 163730: contig of 25916 bp in length
* 163830: gap of unknown length
* 163830: contig of 26266 bp in length
* 190096: gap of unknown length
* 190196: contig of 16099 bp in length
* 206295: gap of unknown length
* 206395: contig of 4405 bp in length
* 210800: gap of unknown length
* 210800: contig of 2715 bp in length
* 210900: gap of unknown length
* 213615: contig of 1066 bp in length.
* 213715: 214780: contig of 1066 bp in length.
*
* Location/Qualifiers
* 1. 214780
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="5"
* /clone="RP23-354F6"
*
* BASE COUNT 68820 a 39755 c 39241 g 66060 t 904 others
* ORIGIN
*
* Query Match 87.0% Score 17.4; DB 2; Length 214780;
* Best Local Similarity 94.7%; Pred. No. 1.2e+02;
* Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
*
* QY 1 attaccacccccctcagg 19
* ||| ||||| ||||| |||||
* Db 163143 ATTTCCACCCCTCGGG 163125
*
* RESULT 13
* AC004126/c
* LOCUS
* DEFINITION Human Chromosome 11q12.2 PAC clone pDJ606g6, complete sequence.
* ACCESSION AC004126
* VERSION AC004126.1 GI:3242744
* KEYWORDS HTG.
* SOURCE human.
* ORGANISM Homo sapiens
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
*
* REFERENCE
* AUTHORS
* 1 (bases 1 to 136328)
* Evans, G.A., Athanasios, M., Aguayo, P., Armstrong, D., Basit, M.,
* Buettner, J., Bumeister, R., Fard, P., deSailbois, F., Dunn, J.,
* English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,
* Grant, O., Hahner, L., Harris, J., Lewis, E., Lou, H., Luo, K.N.,
* Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,
* Schageman, J., Schultz, R.A., Stinson, S., Syed, M. and Ward, T.
* HTGS Submission
* Unpublished
* 2 (bases 1 to 136328)
* Evans, G.A., Athanasios, M., Basit, M., Bradbury, P., Brignac, S.,
* Bumesier, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
* Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
* Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
* Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
*
* TITLE
* JOURNAL
* REFERENCE
* AUTHORS

```

Ward,T. and Wilson,R.
 Direct Submission
 Submitted (06-FEB-1998) Genome Science & Technology Center,
 University of Texas Southwestern Medical Center, 5323 Harry Hines
 Blvd, Dallas, TX 75235-8591, USA

REFERENCE
 AUTHORS 3 (bases 1 to 136328)
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
 Buettner,J., Bumeister,K., Card,P., Desailboat,F., Dunn,J.,
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
 Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
 Schagenan,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
 Direct Submission
 Submitted (20-JUN-1998) Genome Science & Technology Center,
 University of Texas Southwestern Medical Center, 5323 Harry Hines
 Blvd, Dallas, TX 75235-8591, USA

COMMENT
 On Jun 20, 1998 this sequence version replaced gi:2842778.
 IMPORTANT: This submission contains the entire insert of clone
 pDJ60696. pDJ60696 comes from a PAC library constructed at the
 Roswell Park Cancer Institute by the Pieter de Jong group. This
 clone has been finished according to strict quality criteria and
 attempts have been made to resolve all base calling problems such
 as compressions and repetitive elements. The expected Phred/Phrap
 calculated errors/10kb is 0.44. In addition, attempts have been
 made to assure over 98% of consensus base calls consist of either
 double-stranded coverage or 2 types of labeling chemistry on one
 strand.

CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome
 11p12.2 Best's disease region mapped between STS D11S461 and EST
 AHNAK. This region spans over 1.5 Mbp.
 MARKER CONFIRMATION: Sequence confirmed STSs: D11S461, SHGC-36412,
 WI-12296

FEATURES
 MAPED CLONE OVERLAP: HTGS phase 1 PAC clone pDJ1081b4.
 Location/Qualifiers
 1..136328
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 3..111
 /rpt_family="Alu"
 repeat_region 4399..4574
 repeat_region /rpt_family="Alu"
 repeat_region 5447..5700
 repeat_region /rpt_family="Alu"
 repeat_region 6418..6708
 repeat_region /rpt_family="Alu"
 repeat_region 8866..9147
 repeat_region /rpt_family="Alu"
 repeat_region 10302..10604
 repeat_region /rpt_family="Alu"
 repeat_region 10827..11119
 repeat_region /rpt_family="Alu"
 repeat_region complement(12208..12498)
 repeat_region /rpt_family="Alu"
 repeat_region complement(13578..13749)
 repeat_region /rpt_family="MER39"
 repeat_region complement(13647..13737)
 repeat_region /rpt_family="MER21"
 repeat_region complement(13805..13947)
 repeat_region /rpt_family="MER39"
 repeat_region complement(13822..13908)
 repeat_region /rpt_family="MER21"
 repeat_region complement(14283..14759)
 repeat_region /rpt_family="MLT1"
 repeat_region 14798..14988
 repeat_region /rpt_family="MER20"
 repeat_region complement(15469..15753)
 repeat_region /rpt_family="Alu"
 repeat_region complement(16650..16934)
 repeat_region /rpt_family="Alu"
 repeat_region complement(18976..19291)
 repeat_region /rpt_family="Alu"
 repeat_region 19340..19651
 repeat_region /rpt_family="Alu"

repeat_region 20030..20309
 /rpt_family="Alu"
 repeat_region 20780..21041
 /rpt_family="Alu"
 repeat_region complement(23675..23976)
 /rpt_family="Alu"
 repeat_region complement(24041..24140)
 /rpt_family="MIR"
 repeat_region complement(25164..25433)
 /rpt_family="Alu"
 repeat_region 26572..26622
 /rpt_family="MER5"
 repeat_region 27020..27310
 /rpt_family="Alu"
 repeat_region 28214..28492
 /rpt_family="Alu"
 repeat_region complement(30334..30612)
 /rpt_family="Alu"
 repeat_region complement(34144..34257)
 /rpt_family="MIR"
 repeat_region 41000..41307
 /rpt_family="Alu"
 repeat_region complement(42870..42938)
 /rpt_family="MIR"
 repeat_region complement(43280..43389)
 /rpt_family="MIR"
 repeat_region complement(46367..46430)
 /rpt_family="MIR"
 repeat_region complement(47422..47487)
 /rpt_family="MIR"
 repeat_region complement(48093..48170)
 /rpt_family="MIR"
 repeat_region complement(53089..53328)
 /rpt_family="Alu"
 repeat_region complement(56320..56402)
 /rpt_family="MIR"
 repeat_region 56531..56814
 /rpt_family="Alu"
 repeat_region complement(58212..58345)
 /rpt_family="MIR"
 repeat_region 61050..61105
 /rpt_family="LTR1"
 repeat_region 61377..61582
 /rpt_family="LTR1"
 repeat_region 62646..62944
 /rpt_family="Alu"
 repeat_region 64419..64772
 /rpt_family="THE1"
 repeat_region complement(66319..66491)
 /rpt_family="MIR"
 repeat_region 67205..67363
 /rpt_family="Alu"
 repeat_region complement(68725..68847)
 /rpt_family="MIR"
 repeat_region complement(69685..70120)
 /rpt_family="MER21"
 repeat_region 72672..72773
 /rpt_family="MIR"
 repeat_region 77045..77733
 /rpt_family="MER7"
 repeat_region complement(77739..78094)
 /rpt_family="THE1"
 repeat_region 78103..78301
 /rpt_family="MER7"
 repeat_region 78300..80557
 /rpt_family="Trigger"
 repeat_region complement(80554..80788)
 /rpt_family="Alu"
 repeat_region 80809..80882
 /rpt_family="Trigger"
 repeat_region complement(80965..81266)
 /rpt_family="Alu"
 repeat_region 81276..81471

```

/rpt_family="MER7"
84880..85169
/rpt_family="Alu"
complement(85308..85379)
/rpt_family="Alu"
85422..85697
/rpt_family="Alu"
85814..86014
/rpt_family="MIR"
87021..87107
/rpt_family="MIR"
complement(87265..93153)
/rpt_family="L1"
89442..89730
/rpt_family="Alu"
complement(92131..92348)
/rpt_family="MER25"
complement(97690..97823)
/rpt_family="MIR"
106147..106441
/rpt_family="Alu"
108079..108291
/rpt_family="L1"
108358..108646
/rpt_family="Alu"
108668..109058
/rpt_family="L1"
complement(109810..109942)
/rpt_family="MER42"
complement(110662..110778)

Query Match      85.0%  Score 17: DB 9: Length 136328;
Best Local Similarity 100.0%: Pred. No. 2e+02;
Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

QY 4 accacacccctcgga 20
|||||
Db 49357 ACCACCCCTCGGA 49341

RESULT 14
AC087673
LOCUS      AC087673 158991 bp DNA HTG 09-MAY-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-68H20 map 11, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION AC087673
VERSION    AC087673.4 GI:13959181
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 158991)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguski,M., Bouckgeer,H., Brown,A.,
            Cantara,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collumore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Giese,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., Larkocque,K., Lamazares,R., Landers,T.,
            Lenocsky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meidrin,J., Meneus,L., Mihova,T., Mienga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetli,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 7, 2001 this sequence version replaced gi:13940671.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LJ2228
Center clone name: 68_H20
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153011 bases at least Q40
Consensus quality: 155855 bases at least Q30
Consensus quality: 156926 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 157691; sum-of-contigs
Quality coverage: 7.5 in Q20 bases; agarose-fp
Quality coverage: 8.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 613: contig of 613 bp in length
* 614 713: gap of 100 bp
* 714 1405: contig of 692 bp in length
* 1406 1505: gap of 100 bp
* 1506 5561: contig of 4056 bp in length
* 5562 5661: gap of 100 bp
* 5662 8335: contig of 2674 bp in length
* 8336 8435: gap of 100 bp
* 8436 11700: contig of 3265 bp in length
* 11701 11800: gap of 100 bp
* 11801 14386: contig of 2586 bp in length
* 14387 14486: gap of 100 bp
* 14487 46052: contig of 31566 bp in length
* 46053 46152: gap of 100 bp
* 46153 61179: contig of 15027 bp in length
* 61180 61279: gap of 100 bp
* 61280 74186: contig of 12907 bp in length
* 74187 74286: gap of 100 bp
* 74287 90906: contig of 16620 bp in length
* 90907 91006: gap of 100 bp
* 91007 112059: contig of 21053 bp in length
* 112160 112159: gap of 100 bp
* 112160 135292: contig of 23133 bp in length
* 135293 135392: gap of 100 bp
* 135393 156563: contig of 21171 bp in length
* 156564 156663: gap of 100 bp
* 156664 158991: contig of 2328 bp in length.
*
* Location/Qualifiers
* 1..158991
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="11"
* /map="11"
FEATURES
source

```


misc_feature /clone_lib="RP11-68H20"
1. .613
clone_end:SP6
vector_side:left
714. 1405
/note="assembly_fragment"
/note="assembly_fragment"
1506. .5561
/note="assembly_fragment"
5662. .8335
/note="assembly_fragment"
8436. .11700
/note="assembly_fragment"
11801. 14386
/note="assembly_fragment"
14487. .46052
/note="assembly_fragment"
46153. .61179
/note="assembly_fragment"
61280. .74186
/note="assembly_fragment"
74287. .90906
/note="assembly_fragment"
91007. .112059
/note="assembly_fragment"
112160. .135292
/note="assembly_fragment"
135393. .156563
/note="assembly_fragment"
156664. .158991
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 38319 a 39985 c 39117 g 40263 t 1307 others
ORIGIN

COMMENT

On Jan 4, 2001 this sequence version replaced gi:8117601.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft11
Center clone name: RP11-68H20
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.950329
Consensus quality: 157217 bases at least Q40
Consensus quality: 158785 bases at least Q30
Consensus quality: 159537 bases at least Q20
Insert size: 160045; sum-of-contigs
Quality coverage: 9.33x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

25040 contig of 25040 bp in length
1 25141 43548 contig of 18408 bp in length
43649 60068 contig of 16420 bp in length
60169 73827 contig of 13659 bp in length
73928 91951 contig of 18024 bp in length
92052 106957 contig of 14906 bp in length
107058 120803 contig of 13746 bp in length
120904 135414 contig of 14511 bp in length
135515 144584 contig of 9070 bp in length
144685 151176 contig of 6492 bp in length
151277 156342 contig of 5066 bp in length
156443 159750 contig of 3308 bp in length
159851 161245 contig of 1395 bp in length.

* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * as soon as it is available and the accession number will be preserved.

1 25040; contig of 25040 bp in length
* 25041 25140: gap of 100 bp
* 25141 43548: contig of 18408 bp in length
* 43549 43648: gap of 100 bp
* 43649 60068: contig of 16420 bp in length
* 60069 60168: gap of 100 bp
* 60169 73827: contig of 13659 bp in length
* 73828 73927: gap of 100 bp
* 73928 91951: contig of 18024 bp in length
* 91952 92051: gap of 100 bp
* 92052 106957: contig of 14906 bp in length
* 106958 107057: gap of 100 bp
* 107058 120803: contig of 13746 bp in length
* 120804 120903: gap of 100 bp
* 120904 135414: contig of 14511 bp in length
* 135415 135514: gap of 100 bp
* 135515 144584: contig of 9070 bp in length
* 144585 144684: gap of 100 bp
* 144685 151176: contig of 6492 bp in length
* 151177 151276: gap of 100 bp
* 151277 156342: contig of 5066 bp in length
* 156343 156442: gap of 100 bp
* 156443 159750: contig of 3308 bp in length
* 159751 159850: gap of 100 bp
* 159851 161245: contig of 1395 bp in length.

Location/Qualifiers

FEATURES

Query Match 85.0%; Score 17; DB 2; Length 158991;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 accaccacccctcgga 20
|||||
Db 58931 ACCACCACCCCTCGGGA 58947

RESULT 15
AP001986
LOCUS Homo sapiens chromosome 11 clone RP11-68H20 map 11q, WORKING DRAFT
DEFINITION AP001986 161245 bp DNA HTG 04-JAN-2001
SEQUENCE, 13 unordered pieces.
ACCESSION AP001986
VERSION AP001986.3 GI:12039028
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-68H20.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161245)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tolok,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 161,245 genomic DNA of 11q
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 161245)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tolok,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2000) Mashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)

```
source
1..161245
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-68H20"
1..25040
/note="assembly_fragment"
25141..43548
/note="assembly_fragment clone_end:T7 vector_side:right"
43849..60068
/note="assembly_fragment"
60169..73827
/note="assembly_fragment"
73928..91951
/note="assembly_fragment"
92052..106937
/note="assembly_fragment"
107058..120803
/note="assembly_fragment"
120904..135414
/note="assembly_fragment"
135515..144584
/note="assembly_fragment"
144685..151176
/note="assembly_fragment"
151277..156342
/note="assembly_fragment"
156443..159750
/note="assembly_fragment clone_end:SP6 vector_side:right"
159851..161245
/note="assembly_fragment"
BASE COUNT 40274 a 40858 c 39209 g 39704 t 1200 others
ORIGIN

Query Match      85.0%; Score 17; DB 2; Length 161245;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 accacccccctcgga 20
Db 156748 ACCACCCCCCTCGGGA 156764
```

Search completed: February 20, 2002, 11:02:39
Job time: 12067 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:31:58 ; Search time 363.89 Seconds
(without alignments)
47.120 Million cell updates/sec

Title: US-09-904-420A-8
Perfect score: 20
Sequence: 1 attaccaccacccctcgga 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	17	85.0	26555	22	AAI62833
C 2	15.8	79.0	1747	13	AAO31327
C 3	15.8	79.0	1747	13	AAO31643
C 4	15.8	79.0	1749	13	AAO31325
C 5	15.8	79.0	1749	13	AAO31472
C 6	15.8	79.0	1749	13	AAO31629
C 7	15.8	79.0	4049	16	AAQ90660
C 8	15.8	79.0	4097	16	AAQ90657
C 9	15.4	77.0	483	21	AAO80699
C 10	15.4	77.0	875	21	AAI14929
C 11	15.4	77.0	902	22	AAH70729

C 12	15.4	77.0	1867	22	AAH46935	Human secreted pro
C 13	15.4	77.0	1976	22	AAH46935	Human cDNA encodin
C 14	15.4	77.0	2311	22	AAH13994	Human cDNA sequenc
C 15	15.4	77.0	3076	22	AAH46935	Human secreted pro
C 16	15.4	77.0	3370	21	AAH16630	Human secreted pro
C 17	15.4	77.0	8182	22	AAH24879	Nucleotide sequenc
C 18	15.2	76.0	797	22	AAH03404	Human cDNA clone (
C 19	15.2	76.0	1737	22	AAH15253	Human cDNA sequenc
C 20	15.2	76.0	3933	21	AAH37110	Human cDNA sequenc
C 21	15.2	76.0	3934	22	AAH54425	Primer #92 used in
C 22	15	75.0	1209	21	AAH02464	Human colon cancer
C 23	14.8	74.0	343	18	AAH51191	Human breast speci
C 24	14.8	74.0	508	13	AAO31287	Fragment 4, a PCR
C 25	14.8	74.0	508	13	AAO31639	E. brunetti probe
C 26	14.8	74.0	510	21	AAH09338	Fusarium venenatum
C 27	14.8	74.0	526	22	AAH94068	Primer specific fo
C 28	14.8	74.0	723	21	AAH13470	Aspergillus oryzae
C 29	14.8	74.0	1584	22	AAH93866	Human cDNA encodin
C 30	14.8	74.0	1584	22	AAH29359	Anyloid-beta prote
C 31	14.8	74.0	1654	21	AAH59710	Human secreted pro
C 32	14.8	74.0	1702	21	AAH91783	Human Wnt-6 protei
C 33	14.8	74.0	1744	13	AAO31638	Eimeria brunetti s
C 34	14.8	74.0	1744	13	AAO31323	E. brunetti ssRNA
C 35	14.8	74.0	1744	13	AAO31470	E. brunetti ssRNA
C 36	14.8	74.0	1750	13	AAO31324	E. maxima ssRNA
C 37	14.8	74.0	1750	13	AAO31471	E. maxima ssRNA
C 38	14.8	74.0	1750	13	AAO31633	Eimeria maxima ssr
C 39	14.8	74.0	1782	21	AAH21846	Human breast and o
C 40	14.8	74.0	1924	20	AAH80358	R. rubrum poly-bet
C 41	14.8	74.0	3462	22	AAH18415	Human cDNA sequenc
C 42	14.8	74.0	4370	19	AAH32440	p gene promoter Sl
C 43	14.8	74.0	9611	22	AAH86645	Nucleotide sequenc
C 44	14.8	74.0	9611	22	AAH86646	Nucleotide sequenc
C 45	14.8	74.0	9611	22	AAH86647	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAI62833/c
ID AAI62833 standard; DNA; 26555 BP.
XX
AC AAI62833;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 161.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
ds.
XX
CS Homo sapiens.
XX
PN WO200155449-A1.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.

PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 17-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249285.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476225/51.
XX Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
XX Example 2; SEQ ID NO 161; 532pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AA162752-AA162961) and proteins
CC (AAW42347-AAW42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 26555 BP; 5916 A; 7244 C; 7332 G; 6063 T; 0 other;

Query Match 85.0%; Score 17; DB 22; Length 26555;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 accaccaccctcagga 20
DB 25920 ACCCACCCTCCGCGGA 25904

RESULT 2
AAQ31327/C
ID AAQ31327 standard; DNA; 1747 BP.
XX
XX AC
XX AC
XX AAQ31327;

DT 20-APR-1993 (first entry)
XX
DE E. praecox ssrRNA.
XX
KW Small subunit; ribosomal RNA; amplification; PCR; Eimeria; ss.
XX
PN EP516385-A.
XX
XX 02-DEC-1992.
PD
XX 27-MAY-1992; 92EP-0304781.
PF
XX 29-MAY-1991; 91US-0707362.
PR
XX 12-MAY-1992; 92US-0879469.
PR
XX (MERI) MERCK & CO INC.
PA
XX Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;
PI Liberator P, P-Juchelka H;
PI
XX WPI; 1992-400736/49.
DR
XX Species-specific Eimeria tenella DNA probes - comprise divergent
PT DNA sequences and are complementary to E. tenella small sub-unit
PT ribosomal RNA gene
XX
XX Disclosure: Fig 6; 79pp; English.
XX
XX Comparative analysis of regions close to both the 3' and 5' ends
CC of small subunit ribosomal RNA sequences with near identity in the
CC eukaryotic kingdom identified two consensus sequences, ERIR 1 and
CC ERIR 10. ERIR 1 represents a consensus sequence less than 10
CC nucleotides from the 5' end of eukaryotic ssrRNA genes. The ERIR 10
CC sequence is the inverse complement to a consensus sequence located
CC approx. 20 nucleotides from the 3' end of eukaryotic ssrRNA genes.
CC Taken together, these two oligonucleotides span the vast majority of
CC the ssrRNA gene sequence, and they were used as a primer pair in
CC PCR to selectively amplify the ssrRNA genes contained within the
CC genomic DNA prep. from Eimeria praecox. The ssrRNA sequence
CC may then be used to design a species specific probes for diagnosis
CC of Eimeria praecox infection. See also AAQ31283-332.
CC NOTE: As specifications EP-516391, EP-516393-6, EP-516391 and
CC EP-516395-6 are identical except in the claims section, sequences
CC for all these specifications can be found indexed under EP-516385.
CC However the claimed sequences of each specification will be indexed
CC under their own patent number, thus each separate patent will be
CC represented.
XX
XX Sequence 1747 BP; 445 A; 347 C; 471 G; 484 T; 0 other;

Query Match 79.0%; Score 15.8; DB 13; Length 1747;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccaccctcggg 19
DB 1502 ATTACCACCCCTCGGG 1484

RESULT 3
AAQ31643/C
ID AAQ31643 standard; DNA; 1747 BP.
XX
XX AC
XX AAQ31643;
XX
XX 20-APR-1993 (first entry)
DE Eimeria praecox ssrRNA.
XX
XX Small subunit; ribosomal RNA; amplification; PCR; ss.
XX
XX EP516384-A.

```
XX PD 02-DEC-1992.
XX PF 27-MAY-1992; 92EP-0304780.
XX PR 29-MAY-1991; 91US-0707360.
XX PR 12-MAY-1992; 92US-0879594.
XX PA (MERI ) MERCK & CO INC.
XX PI Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;
XX PI Liberator P, P-Juchelka H;
XX DR WPI; 1992-400735/49.
XX PT Species-specific Eimeria tenella DNA probes - comprise divergent
XX PT DNA sequences and are complementary to E. tenella small sub-unit
XX PT ribosomal RNA gene
XX PA Claim 10; Fig 6; 9lpp; English.
XX PS Comparative analysis of regions close to both the 3' and 5' ends
XX CC of small subunit ribosomal RNA sequences with near identity in the
XX CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and
XX CC ERIB 10. Taken together, these two oligonucleotides span the vast
XX CC majority of the ssrRNA sequence, and they were used as a primer pair
XX CC in PCR to selectively amplify the ssrRNA gene contained within genomic
XX CC DNA prep. from Eimeria praecox. The ssrRNA sequence may then be used
XX CC to design species specific probes for diagnosis of Eimeria praecox
XX CC infection, e.g. regions 106-113, 153-179, 188-223, 253-266,
XX CC 630-727, 927-933, 1031-1046, 1056-1062, 1154-1168, 1335-1375, 1471-
XX CC 1514 and 1664-1699 (claimed) which are complementary to the ssrRNA
XX CC gene of E. praecox. See also AAO31640-2.
XX CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and
XX CC EP-516395-6 are identical except in the claims section, sequences
XX CC for all these specifications can be found indexed under EP-516385.
XX CC However the claimed sequences of each specification will be indexed
XX CC under their own patent number, thus each separate patent will be
XX CC represented.
XX SQ Sequence 1747 BP; 446 A; 348 C; 471 G; 482 T; 0 other;

Query Match 79.0%; Score 15.8; DB 13; Length 1747;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
DB 1502 ATTACCCAGCCCATCGGG 1484

RESULT 4
AAO31325/c
ID AAO31325 standard; DNA: 1749 BP.
XX AC AAO31325;
XX DT 20-APR-1993 (first entry)
XX DE E. mitis ssrRNA.
XX KW Small subunit; ribosomal RNA; amplification; PCR; Eimeria; ss.
XX PN EP516385-A.
XX PD 02-DEC-1992.
XX PF 27-MAY-1992; 92EP-0304781.
XX PR 29-MAY-1991; 91US-0707362.
XX PR 12-MAY-1992; 92US-0879469.
XX

PA (MERI ) MERCK & CO INC.
XX Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;
XX PI Liberator P, P-Juchelka H;
XX DR WPI; 1992-400735/49.
XX PT Species-specific Eimeria tenella DNA probes - comprise divergent
XX PT DNA sequences and are complementary to E. tenella small sub-unit
XX PT ribosomal RNA gene
XX PS Disclosure; Fig 4; 79pp; English.
XX CC Comparative analysis of regions close to both the 3' and 5' ends
XX CC of small subunit ribosomal RNA sequences with near identity in the
XX CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and
XX CC ERIB 10. ERIB 1 represents a consensus sequence less than 10
XX CC nucleotides from the 5' end of eukaryotic ssrRNA genes. The ERIB 10
XX CC sequence is the inverse complement to a consensus sequence located
XX CC approx. 20 nucleotides from the 3' end of eukaryotic ssrRNA genes.
XX CC Taken together, these two oligonucleotides span the vast majority of
XX CC the ssrRNA gene sequence, and they were used as a primer pair in
XX CC PCR to selectively amplify the ssrRNA genes contained within the
XX CC genomic DNA prep. from Eimeria mitis. The ssrRNA sequence
XX CC may then be used to design a species specific probes for diagnosis
XX CC of Eimeria mitis infection. See also AAO31283-332.
XX CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and
XX CC EP-516395-6 are identical except in the claims section, sequences
XX CC for all these specifications can be found indexed under EP-516385.
XX CC However the claimed sequences of each specification will be indexed
XX CC under their own patent number, thus each separate patent will be
XX CC represented.
XX SQ Sequence 1749 BP; 437 A; 362 C; 465 G; 485 T; 0 other;

Query Match 79.0%; Score 15.8; DB 13; Length 1749;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
DB 1505 ATTACCCAGCCCATCGGG 1487

RESULT 5
AAO31472/c
ID AAO31472 standard; DNA: 1749 BP.
XX AC AAO31472;
XX DT 20-APR-1993 (first entry)
XX DE E. mitis ssrRNA.
XX KW Small subunit; ribosomal RNA; amplification; PCR; Eimeria; ss.
XX PN EP516381-A.
XX PD 02-DEC-1992.
XX PF 27-MAY-1992; 92EP-0304777.
XX PR 29-MAY-1991; 91US-0707356.
XX PR 12-MAY-1992; 92US-0879585.
XX PA (MERI ) MERCK & CO INC.
XX PI Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;
XX PI Liberator P, P-Juchelka H;
XX DR WPI; 1992-400732/49.
XX
```

PT Species-specific *Elmeria tenella* DNA probes - comprise divergent
 PT DNA sequences and are complementary to *E. tenella* small sub-unit
 PT ribosomal RNA gene
 XX
 PS Disclosure; Fig 4; 79pp; English.
 XX
 CC Comparative analysis of regions close to both the 3' and 5' ends
 CC of small subunit ribosomal RNA sequences with near identity in the
 CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and
 CC ERIB 10. ERIB 1 represents a consensus sequence less than 10
 CC nucleotides from the 5' end of eukaryotic ssrRNA genes. The ERIB 10
 CC sequence is the inverse complement to a consensus sequence located
 CC approx. 20 nucleotides from the 3' end of eukaryotic ssrRNA genes.
 CC Taken together, these two oligonucleotides span the vast majority of
 CC the ssrRNA gene sequence, and they were used as a primer pair in
 CC PCR to selectively amplify the ssrRNA genes contained within the
 CC genomic DNA prep. from *Elmeria mitis*. The ssrRNA sequence
 CC may then be used to design a species specific probes for diagnosis
 CC of *Elmeria mitis* infection (method claimed). See also AAQ31465-71.
 CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and
 CC EP-516395-6 are identical except in the claims section, sequences
 CC for all these specifications can be found indexed under EP-516385.
 CC However the claimed sequences of each specification will be indexed
 CC under their own patent number, thus each separate patent will be
 CC represented.
 XX
 SQ Sequence 1749 BP; 437 A; 362 C; 465 G; 485 T; 0 other;

Query Match 79.0%; Score 15.8; DB 13; Length 1749;
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccacccccctcggg 19
 ||||| ||| |||||
 DB 1505 ATTACCCAGCCCATCGGG 1487

RESULT 6
 AAQ31629/c
 ID AAQ31629 standard; DNA; 1749 BP.

AC AAQ31629;
 DT 20-APR-1993 (first entry)
 XX
 DE *Elmeria mitis* ssrRNA.
 XX
 KW Small subunit; ribosomal RNA; amplification; PCR; ss.
 XX
 PN EP516386-A.
 XX
 PD 02-DEC-1992.

PF 27-MAY-1992; 92EP-0304782.
 XX
 XX 29-MAY-1991; 91US-0707355.
 PR 12-MAY-1992; 92US-0879640.

XX
 PA (MERI) MERCK & CO INC.
 XX
 PI Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;
 PI Liberator P, P-Juchelka H;
 XX
 DR WPI; 1992-400737/49.

XX Species-specific *Elmeria tenella* DNA probes - comprise divergent
 PT DNA sequences and are complementary to *E. tenella* small sub-unit
 PT ribosomal RNA gene
 XX
 PS Claim 10; Fig 4; 79pp; English.
 XX
 CC Comparative analysis of regions close to both the 3' and 5' ends

CC of small subunit ribosomal RNA sequences with near identity in the
 CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and
 CC ERIB 10. Taken together, these two oligonucleotides span the vast
 CC majority of the ssrRNA sequence, and they were used as a primer pair
 CC in PCR to selectively amplify the ssrRNA gene contained within genomic
 CC DNA prep. from *Elmeria mitis*. The ssrRNA sequence may then be used
 CC to design species specific probes for diagnosis of *Elmeria mitis*
 CC infection, e.g. regions 106-113, 153-179, 188-223, 253-266,
 CC 630-725, 923-928, 1026-1041, 1051-1057, 1149-1161, 1329-1380, 1474-
 CC 1517 and 1667-1701 (claimed) which are complementary to the ssrRNA
 CC gene of *E. mitis*. See also AAQ31626-8.
 CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and
 CC EP-516395-6 are identical except in the claims section, sequences
 CC for all these specifications can be found indexed under EP-516385.
 CC However the claimed sequences of each specification will be indexed
 CC under their own patent number, thus each separate patent will be
 CC represented.
 XX
 SQ Sequence 1749 BP; 437 A; 362 C; 465 G; 485 T; 0 other;

Query Match 79.0%; Score 15.8; DB 13; Length 1749;
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccacccccctcggg 19
 ||||| ||| |||||
 DB 1505 ATTACCCAGCCCATCGGG 1487

RESULT 7
 AAQ90660/c
 ID AAQ90660 standard; DNA; 4049 BP.

AC AAQ90660;
 XX
 DT 11-NOV-1995 (first entry)
 XX
 DE Eph-related PTK Cdk5.
 XX
 KW Cdk5; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;
 KW prognosis; ss.
 XX
 OS *Gallus* sp.

XX
 FH Key Location/Qualifiers
 FT CDS 10..2997
 FT /*tag= a

XX WO9515375-A.
 XX
 PD 08-JUN-1995.
 XX
 PF 07-SEP-1994; 94WO-US10140.
 XX
 PR 03-DEC-1993; 93US-0162809.
 XX
 PA (LJOI-) LA JOLLA CANCER RES FOUND.

XX .Pasquale EB, Sajjadi FG;
 XX
 XX WPI; 1995-215256/28.
 DR P-PSDB; AAR75712.

XX Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing
 PT cancer.
 XX
 PS Disclosure; Page 92-96; 129pp; English.

XX Probes derived from the EPH-related PTKs Cdk5 (AAQ90659) and Cdk5
 CC (AAQ90660) were used to isolate novel cDNA clones (AAQ90652-58,
 CC AAQ90661-62) from chicken embryo and embryonic brain libraries.
 CC

SQ Sequence 4049 BP; 1010 A; 1011 C; 1072 G; 956 T; 0 other;

Query Match 79.0%; Score 15.8; DB 16; Length 4049;
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
 ||| ||||| ||||| |||||
 Db 3686 ATTCCCCACCCCCCGGG 3668

RESULT 8
 AAQ0657/C
 ID AAQ0657 standard; DNA; 4097 BP.
 XX AC AAQ0657;
 XX DT 11-NOV-1995 (first entry)
 XX DE Eph-related PTK Cdk5+.
 XX KW Cdk5+; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;
 XX KW prognosis; ss.
 XX OS Gallus sp.
 XX FH Key Location/Qualifiers
 FT CDS 10..3045
 FT /*tag= a
 XX PN W09515375-A.
 XX PD 08-JUN-1995.
 XX PF 07-SEP-1994; 94WO-US10140.
 XX PR 03-DEC-1993; 93US-0162809.
 XX PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX PI Pasquale EB, Sajjadi FG;
 XX WP1; 1995-215256/28.
 XX DR P-PSDB; AAR75709.
 XX PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing
 XX PS Cancer.
 XX PS Claim 2; Page 71-75; 129pp; English.
 CC A cDNA clone encoding a novel variant of Eph-related PTK Cdk5,
 CC Cdk5+ (AAQ0657), was isolated from a chick embryo library in
 CC lambda gt11. Cdk5+ protein (AAR75709) contains a 16-amino acid
 CC insertion in the juxtamembrane domain, and be a result of
 CC alternative splicing. Cdk5+ is exclusively expressed in the CNS.
 XX Sequence 4097 BP; 1014 A; 1029 C; 1087 G; 967 T; 0 other;

Query Match 79.0%; Score 15.8; DB 16; Length 4097;
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
 ||| ||||| ||||| |||||
 Db 3734 ATTCCCCACCCCCCGGG 3716

RESULT 9
 AAC08069/c
 ID AAC08069 standard; cDNA; 483 BP.
 XX

AAC08069;
 06-OCT-2000 (first entry)
 Human secreted protein 5' EST, SEQ ID NO: 12144.
 Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 gene therapy; chromosome mapping; ss.
 Homo sapiens.
 EP1033401-A2.
 06-SEP-2000.
 21-FEB-2000; 2000EP-0200610.
 26-FEB-1999; 99US-0122487.
 (GEST) GENSET.
 Dumas Milne Edwards J, Duclert A, Giordano J;
 WPI; 2000-500381/45.
 New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures -
 Claim 1; SEQ ID 12144; 71pp + CD-ROM; English.
 The present sequence is one of a large number of 5' ESTs derived from
 mRNAs encoding secreted proteins. No ORF has yet been conclusively
 identified within the present sequence. The 5' ESTs were prepared from
 total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 sequences usually correspond mainly to the 3' untranslated region (UTR)
 of the mRNA because they are often obtained from oligo-dT primed cDNA
 libraries. Such ESTs are not well suited for isolating cDNA sequences
 derived from the 5' ends of mRNAs and even in those cases where longer
 cDNA sequences have been obtained, the full 5' UTR is rarely included.
 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 They are used to obtain upstream regulatory sequences and to design
 expression and secretion vectors.
 Sequence 483 BP; 142 A; 63 C; 96 G; 180 T; 2 other;

Query Match 77.0%; Score 15.4; DB 21; Length 483;
 Best Local Similarity 94.1%; Pred. No. 2.4e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attaccacccccctcgg 17
 ||||| ||||| ||||| |||||
 Db 145 ATTACCCACCCCCCTTG 129

RESULT 10
 AAF14929
 ID AAF14929 standard; cDNA; 875 BP.
 XX AC AAF14929;
 XX DT 13-MAR-2001 (first entry)
 XX DE Trichoderma reesei EST SEQ ID NO:7452.
 XX KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.


```
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
XX PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
XX PI Ni J, Ruben SM, Barash SC;
XX DR WPI: 2001-476220/51.
XX DR P-PSDB; AAB85525.
XX PT 17 isolated nucleic acid molecules encoding human secreted proteins,
XX PT used to preventing, treating or ameliorating a medical condition -
XX PS Claim 1; Page 414-415; 482pp; English.
XX CC The invention provides novel human secreted proteins and polynucleotides
XX CC encoding them. The secreted proteins can be expressed by standard
XX CC recombinant methodology. The secreted proteins and polynucleotides are
XX CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
XX CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
XX CC also be used in diagnosing a pathological condition. The antibodies to
XX CC the proteins can also be used in alleviating symptoms associated with the
XX CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
XX CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
XX CC treated include autoimmune diseases e.g. rheumatoid arthritis,
XX CC hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders
XX CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX CC ocular disorders e.g. corneal infection. The polypeptides can also be
XX CC used to aid wound healing and epithelial cell proliferation, to prevent
XX CC skin aging due to sunburn, to maintain organs before transplantation, for
XX CC supporting cell culture of primary tissues, to regenerate tissues and in
XX CC chemotaxis. The polypeptides can also be used as a food additive or
XX CC preservative to increase or decrease storage capabilities. The present
XX CC sequence represents a human secreted protein encoding cDNA.
XX SQ Sequence 1867 BP; 542 A; 315 C; 347 G; 663 T; 0 Other;

Query Match 77.0%; Score 15.4; DB 22; Length 1867;
Best Local Similarity 94.1%; Pred. No. 2.4e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 attaccacccccctcg 17
Db 1501 ATTACCCACCCCTTG 1485

RESULT 13
AAF93822/c
ID AAF93822 standard; cDNA; 1976 BP.
XX AC AAF93822;
XX DT 23-MAY-2001 (first entry)
XX DE Human cDNA encoding a membrane or secretory protein clone PSEC0143.
XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;
XX KW rheumatoid arthritis; diabetes; ss.
XX OS Homo sapiens.
XX PN EP1067182-A2.
XX PD 10-JAN-2001.
XX PF 07-JUL-2000; 2000EP-0114090.
XX PR 08-JUL-1999; 99JP-0194179.
XX PR 11-JAN-2000; 2000JP-0118775.
XX PR 02-MAY-2000; 2000JP-0183766.
```

```
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX DR WPI: 2001-093989/11.
XX DR P-PSDB; AAB88395.
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX PT gene therapy or as candidate target molecules in drug development -
XX PS Claim 1; SEQ ID 157; 609pp + CD ROM; English.
XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX CC which encode human secretory or membrane proteins represented by
XX CC AAB88317 - AAB88419. Included in the invention are primers
XX CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
XX CC cDNA sequences of the invention. The invention also includes methods for
XX CC the production of antibodies directed against the proteins, and cDNA
XX CC sequences, which can be used in vaccines. The polynucleotide sequences
XX CC can be used in gene therapy. The polynucleotide sequences and the
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with inappropriate secretory
XX CC protein/membrane protein expression. The nucleic acids and complementary
XX CC sequences may also be used as DNA probes in diagnostic assays
XX CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
XX CC presence of similar nucleic acid sequences in samples. They may also be
XX CC used to study the expression and function of secretory proteins/membrane
XX CC polypeptides and their role in metabolism. The polypeptides may be used
XX CC as antigens in the production of antibodies against them and in assays to
XX CC identify modulators (agonists and antagonists) of expression and
XX CC activity. The antibodies and antagonists may also be used as therapeutic
XX CC agents to down regulate expression and activity. The antibodies may also
XX CC be used as diagnostic agents for detecting the presence of the
XX CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
XX CC (ELISA)). Examples of diseases which may be treated include rheumatoid
XX CC arthritis and diabetes.
XX SQ Sequence 1976 BP; 595 A; 314 C; 371 G; 696 T; 0 Other;

Query Match 77.0%; Score 15.4; DB 22; Length 1976;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 attaccacccccctcg 17
Db 1312 ATTACCCACCCCTTG 1296

RESULT 14
AAH13994/c
ID AAH13994 standard; cDNA; 2311 BP.
XX AC AAH13994;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:11070.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX KW Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
```

PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8: SEQ ID 11070; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 2311 BP; 655 A; 382 C; 458 G; 816 T; 0 other;
SQ

Query Match 77.0%; Score 15.4; DB 22; Length 2311;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 attaccacccccctcg 17
Db 1660 ATTACCCACCCCTTG 1944

RESULT 15
AAH46951/C
ID AAH46951 standard; cDNA: 3076 BP.
XX
AC AAH46951;
XX
XX 25-SEP-2001 (first entry)
XX
XX Human secreted protein encoding cDNA (clone Id HB0EG11).
XX
XX Secreted protein; immunosuppressive; antirheumatic; antirheumatic;
KW antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;
KW neotropic; neuroprotective; antibacterial; virucide; fungicide; human;
KW opthalmological; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200155430-A1.
XX
XX 02-AUG-2001.
XX

PF 17-JAN-2001; 2001WO-US01431.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 12-SEP-2000; 2000US-0231968.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI Ni J, Ruben SM, Barash SC;
XX
XX WPI: 2001-476220/51.
DR P-PSDB: AAB85541.
XX
XX 17 isolated nucleic acid molecules encoding human secreted proteins,
PT used to preventing, treating or ameliorating a medical condition -
PT
XX Claim 1: Page 426-427; 482pp; English.
PS
XX The invention provides novel human secreted proteins and polynucleotides
CC encoding them. The secreted proteins can be expressed by standard
CC recombinant methodology. The secreted proteins and polynucleotides are
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
CC also be used in diagnosing a pathological condition. The antibodies to
CC the proteins can also be used in alleviating symptoms associated with the
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
CC treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiogenesis caused by bacteria, viruses and fungi and
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a human secreted protein encoding cDNA.
XX
XX Sequence 3076 BP; 940 A; 487 C; 617 G; 1028 T; 4 other;
SQ

Query Match 77.0%; Score 15.4; DB 22; Length 3076;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 attaccacccccctcg 17
Db 2651 ATTACCCACCCCTTG 2635

Search completed: February 20, 2002, 10:32:01
Job time: 10229 sec

Thu Feb 21 07:26:04 2002

us-09-904-420a-8.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:44:49 ; Search time 166.82 seconds
(without alignments)
27.152 Million cell updates/sec

Title: US-09-904-420A-8
Sequence: 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.8	79.0	1747	1	US-07-879-647A-29
C 2	15.8	79.0	1747	1	US-07-879-584A-29
C 3	15.8	79.0	1747	1	US-07-879-470A-29
C 4	15.8	79.0	1747	1	US-07-879-644A-29
C 5	15.8	79.0	1747	1	US-07-879-640A-29
C 6	15.8	79.0	1747	1	US-07-879-594A-29
C 7	15.8	79.0	1747	1	US-07-879-469A-29
C 8	15.8	79.0	1749	1	US-07-879-647A-27
C 9	15.8	79.0	1749	1	US-07-879-584A-27
C 10	15.8	79.0	1749	1	US-07-879-470A-27
C 11	15.8	79.0	1749	1	US-07-879-644A-27
C 12	15.8	79.0	1749	1	US-07-879-640A-27
C 13	15.8	79.0	1749	1	US-07-879-594A-27
C 14	15.8	79.0	1749	1	US-07-879-469A-27
C 15	15.8	79.0	4049	1	US-08-162-809-17
C 16	15.8	79.0	4097	1	US-08-162-809-11
C 17	14.8	74.0	343	5	PCT-US95-08295-13
C 18	14.8	74.0	508	1	US-07-879-647A-50
C 19	14.8	74.0	508	1	US-07-879-584A-50
C 20	14.8	74.0	508	1	US-07-879-470A-50
C 21	14.8	74.0	508	1	US-07-879-644A-50
C 22	14.8	74.0	508	1	US-07-879-640A-50
C 23	14.8	74.0	508	1	US-07-879-594A-50
C 24	14.8	74.0	508	1	US-07-879-469A-50
C 25	14.8	74.0	1744	1	US-07-879-647A-25
C 26	14.8	74.0	1744	1	US-07-879-584A-25
C 27	14.8	74.0	1744	1	US-07-879-470A-25

C 28	14.8	74.0	1744	1	US-07-879-644A-25
C 29	14.8	74.0	1744	1	US-07-879-640A-25
C 30	14.8	74.0	1744	1	US-07-879-594A-25
C 31	14.8	74.0	1744	1	US-07-879-469A-25
C 32	14.8	74.0	1748	1	US-07-879-647A-24
C 33	14.8	74.0	1748	1	US-07-879-584A-24
C 34	14.8	74.0	1748	1	US-07-879-470A-24
C 35	14.8	74.0	1748	1	US-07-879-644A-24
C 36	14.8	74.0	1748	1	US-07-879-640A-24
C 37	14.8	74.0	1748	1	US-07-879-594A-24
C 38	14.8	74.0	1750	1	US-07-879-469A-24
C 39	14.8	74.0	1750	1	US-07-879-647A-26
C 40	14.8	74.0	1750	1	US-07-879-584A-26
C 41	14.8	74.0	1750	1	US-07-879-470A-26
C 42	14.8	74.0	1750	1	US-07-879-644A-26
C 43	14.8	74.0	1750	1	US-07-879-640A-26
C 44	14.8	74.0	1750	1	US-07-879-594A-26
C 45	14.8	74.0	1750	1	US-07-879-469A-26

ALIGNMENTS

RESULT 1
US-07-879-647A-29/c
; Sequence 29, Application US/07879647A
; Patent No. 5266689
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevich, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Elmeria Maxima DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,647A
; FILING DATE: 19920512
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,628
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: 184201A
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1747 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-879-647A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccacccccctcggg 19

Db 1502 ATTACCAGCCCCATCGGG 1484

RESULT 2

US-07-879-584A-29/c
Sequence 29, Application US/07879584A

Patent No. 5278298

GENERAL INFORMATION:

APPLICANT: Chakraborty, P.R.

APPLICANT: Dashkevicz, M.

APPLICANT: Elbrecht, A.

APPLICANT: Feighner, S.D.

APPLICANT: Liberator, P.A.

APPLICANT: Profous-Juchelka, H.

TITLE OF INVENTION: Eimeria Brunetti DNA

NUMBER OF INVENTION: Probes

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.

STREET: 126 Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: USA

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb

MEDIUM TYPE: storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 6.0.4

SOFTWARE: Microsoft Word 4.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/879,584A

FILING DATE: 19920512

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/706,717

FILING DATE: 29-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Tribble, Jack L.

REGISTRATION NUMBER: 32,633

REFERENCE/DOCKET NUMBER: 184191A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 594-5321

TELEFAX: (908) 594-4720

TELEX: 138825

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 1747 bases

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

US-07-879-584A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccacccccctcggg 19

Db 1502 ATTACCAGCCCCATCGGG 1484

RESULT 3

US-07-879-470A-29/c

Sequence 29, Application US/07879470A

Patent No. 5288845
GENERAL INFORMATION:
APPLICANT: Chakraborty, P.R.
APPLICANT: Dashkevicz, M.
APPLICANT: Elbrecht, A.
APPLICANT: Feighner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Necatrix DNA
NUMBER OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,470A
FILING DATE: 19920512
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,351
FILING DATE: 29-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: 184221A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1747 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-470A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 attaccacccccctcggg 19

Db 1502 ATTACCAGCCCCATCGGG 1484

RESULT 4

US-07-879-644A-29/c

Sequence 29, Application US/07879644A

Patent No. 5298613

GENERAL INFORMATION:

APPLICANT: Chakraborty, P.R.

APPLICANT: Dashkevicz, M.

APPLICANT: Elbrecht, A.

APPLICANT: Feighner, S.D.

APPLICANT: Liberator, P.A.

APPLICANT: Profous-Juchelka, H.

TITLE OF INVENTION: Eimeria Acervulina DNA

NUMBER OF INVENTION: Probes

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.

```

; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette,3.50 inch,800 kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,644A
; FILING DATE: 19920512
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,817
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184218IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1747 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-879-640A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
Db 1502 ATTACCCAGCCCATCGGG 1484

RESULT 6
US-07-879-640A-29/c
; Sequence 29, Application US/07879594A
; Patent No. 5449768
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Praecox DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette,3.50 inch,800 kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,594A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,360
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184231A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
```

```

; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette,3.50 inch,800 kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,644A
; FILING DATE: 19920512
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,817
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184181A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1747 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-879-640A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
Db 1502 ATTACCCAGCCCATCGGG 1484

RESULT 5
US-07-879-640A-29/c
; Sequence 29, Application US/07879640A
; Patent No. 5359050
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Mitis DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette,3.50 inch,800 kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,640A
```

INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1747 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-594A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcg 19
||||| ||| ||| |||
DB 1502 ATTACCCAGCCCATCGG 1484

RESULT 7
US-07-879-469A-29/c
; Sequence 29, Application US/07879469A
; Patent No. 5563256
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Tenella DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,469A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,362
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184241A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1747 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-469A-29

Query Match 79.0%; Score 15.8; DB 1; Length 1747;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcg 19
||||| ||| ||| |||
DB 1502 ATTACCCAGCCCATCGG 1484
RESULT 8
US-07-879-647A-27/c
; Sequence 27, Application US/07879647A
; Patent No. 5266689
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Maxima DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,647A
; FILING DATE: 19920512
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,628
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184201A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-647A-27

Query Match 79.0%; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcg 19
||||| ||| ||| |||
DB 1505 ATTACCCAGCCCATCGG 1487

RESULT 9
US-07-879-584A-27/c
; Sequence 27, Application US/07879584A
; Patent No. 5278298
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.

Thu Feb 21 07:26:04 2002

us-09-904-420a-8.rni

APPLICANT: Feighner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Brunetti DNA
TITLE OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
FILING DATE: 19920512
APPLICATION NUMBER: US/07/879,584A
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,717
FILING DATE: 29-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: .184191A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-584A-27

Query Match 79.08; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
||||| ||| ||| |||
Db 1505 ATTACCACGCCATCGGG 1487

RESULT 10
US-07-879-470A-27/c
Sequence 27, Application US/07879470A
Patent No. 5288845
GENERAL INFORMATION:
APPLICANT: Chakraborty, P.R.
APPLICANT: Dashkevicz, M.
APPLICANT: Elbrecht, A.
APPLICANT: Feighner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Necatrix DNA
TITLE OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,470A
FILING DATE: 19920512
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,351
FILING DATE: 29-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: .184221A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-470A-27

Query Match 79.08; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
||||| ||| ||| |||
Db 1505 ATTACCACGCCATCGGG 1487

RESULT 11
US-07-879-644A-27/c
Sequence 27, Application US/07879644A
Patent No. 5298613
GENERAL INFORMATION:
APPLICANT: Chakraborty, P.R.
APPLICANT: Dashkevicz, M.
APPLICANT: Elbrecht, A.
APPLICANT: Feighner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Acervulina DNA
TITLE OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,644A
FILING DATE: 19920512
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,817
FILING DATE: 29-MAY-1991

ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: .184181A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-644A-27

Query Match 79.0%; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attaccacccccctcggg 19
||||| ||| |||||

Db 1505 ATTACCCAGCCCATCGGG 1487

RESULT 12
US-07-879-640A-27/c
Sequence 27, Application US/07879640A
Patent No. 5359050
GENERAL INFORMATION:
APPLICANT: Chakraborty, P.R.
APPLICANT: Dashkevicz, M.
APPLICANT: Elbrecht, A.
APPLICANT: Felghner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Mitis DNA
TITLE OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,640A
FILING DATE: 19920512
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,355
FILING DATE: 29-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: .184211A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single

TOPOLOGY: linear
US-07-879-640A-27

Query Match 79.0%; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attaccacccccctcggg 19
||||| ||| |||||

Db 1505 ATTACCCAGCCCATCGGG 1487

RESULT 13
US-07-879-594A-27/c
Sequence 27, Application US/07879594A
Patent No. 5449768
GENERAL INFORMATION:
APPLICANT: Chakraborty, P.R.
APPLICANT: Dashkevicz, M.
APPLICANT: Elbrecht, A.
APPLICANT: Felghner, S.D.
APPLICANT: Liberator, P.A.
APPLICANT: Profous-Juchelka, H.
TITLE OF INVENTION: Eimeria Praecox DNA
TITLE OF INVENTION: Probes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
MEDIUM TYPE: storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.4
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,594A
FILING DATE: 19920512
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/706,360
FILING DATE: 29-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tribble, Jack L.
REGISTRATION NUMBER: 32,633
REFERENCE/DOCKET NUMBER: .184231A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-5321
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-879-594A-27

Query Match 79.0%; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attaccacccccctcggg 19
||||| ||| |||||

Db 1505 ATTACCCAGCCCATCGGG 1487

RESULT 14
US-07-879-469A-27/c
; Sequence 27, Application US/07879469A
; Patent No. 5563256
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevich, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Tenella DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,469A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,362
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184241A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-469A-27

Query Match 79.0%; Score 15.8; DB 1; Length 1749;
Best Local Similarity 89.5%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
||||| ||| ||| |||

Db 1505 ATTACCCACGCCCATCGGG 1487

RESULT 15
US-08-162-809-17/c
; Sequence 17, Application US/08162809
; Patent No. 5457048
; GENERAL INFORMATION:
; APPLICANT: Pasquale, Elena B.
; APPLICANT: Sajjadi, Fereydoun G.
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,809
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4049 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; FEATURE: CDS
; NAME/KEY: CDS
; LOCATION: 10..2994
US-08-162-809-17

Query Match 79.0%; Score 15.8; DB 1; Length 4049;
Best Local Similarity 89.5%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcggg 19
||||| ||| ||| |||

Db 3686 ATTCCCCACGCCCATCGGG 3668

Search completed: February 20, 2002, 07:44:50
Job time: 198 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:25:45 ; Search time 3814.84 Seconds
(without alignments)
56.337 Million cell updates/sec

Title: US-09-904-420a-8
Perfect score: 20
Sequence: 1 attaccaccacctcgga 20

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estfun:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estom:*
 - 5: em_estpl:*
 - 6: em_estba:*
 - 7: em_estro:*
 - 8: em_estov:*
 - 9: em_hic:*
 - 10: gb_estl:*
 - 11: gb_est2:*
 - 12: gb_hic:*
 - 13: gb_gss:*
 - 14: em_gss_fun:*
 - 15: em_gss_hum:*
 - 16: em_gss_inv:*
 - 17: em_gss_pln:*
 - 18: em_gss_pro:*
 - 19: em_gss_rod:*
 - 20: em_gss_vrt:*
 - 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20	100.0	353	11	BG386035	BG386035 602455204
2	18	90.0	273	10	AA311787	AA311787 EST182503
3	17.4	87.0	435	10	AW588384	AW588384 pal2803.y
4	17.4	87.0	1735	11	BE965843	BE965843 601659019
5	17	85.0	350	11	BE970922	BE970922 601680096
6	17	85.0	866	11	BF338368	BF338368 602034011
7	16.8	84.0	489	13	A2194055	A2194055 SP_1025.B
8	16.8	84.0	542	10	BE684213	BE684213 184527 WA
9	16.8	84.0	890	10	AW155439	AW155439 maie00170
10	16.8	84.0	896	11	BF106598	BF106598 601823238
11	16.8	84.0	969	10	BE216883	BE216883 HV_CEB001
12	16.4	82.0	281	11	BF964059	BF964059 PM4-NNI20

C 13	16.4	82.0	367	10	AA165834	AA165834 ms64d04.r
C 14	16.4	82.0	368	10	AT640019	AT640019 ms64d04.y
C 15	16.4	82.0	448	10	AA954701	AA954701 oc84a06.s
C 16	16.4	82.0	494	11	BF623736	BF623736 HVSMEa000
C 17	16.4	82.0	617	13	AQ744329	AQ744329 HS_5502_A
C 18	16.4	82.0	797	11	BG344434	BG344434 HVSMEg000
C 19	16.4	82.0	905	11	BF205344	BF205344 601867918
C 20	16.4	82.0	940	10	BE560757	BE560757 601347075
C 21	16.4	82.0	945	11	BF576543	BF576543 602133981
C 22	16.4	82.0	1144	10	BE213792	BE213792 HV_CEB000
C 23	16	80.0	815	13	CNS02JWQ	AL200771 Tetraodon
C 24	16	80.0	897	11	BF973134	BF973134 602242122
C 25	16	80.0	903	13	CNS03WDN	AL263588 Tetraodon
C 26	15.8	79.0	52	10	A1443583	A1443583 sa33h12.x
C 27	15.8	79.0	195	10	BB542990	BB542990 BB542990
C 28	15.8	79.0	211	10	BE708017	BE708017 MR0-HT055
C 29	15.8	79.0	221	10	BB334119	BB334119 RB334119
C 30	15.8	79.0	231	10	BB340755	BB340755 BB340755
C 31	15.8	79.0	242	10	BB262777	BB262777 BB262777
C 32	15.8	79.0	257	10	AV300572	AV300572 AB300572
C 33	15.8	79.0	271	10	A1837244	A1837244 UI-M-AK0-
C 34	15.8	79.0	287	10	AW529603	AW529603 UI-R-BT1-
C 35	15.8	79.0	295	11	BI428911	BI428911 fr69c02.x
C 36	15.8	79.0	298	13	A2514862	A2514862 LM0361D20
C 37	15.8	79.0	300	13	A242345	A242345 RPCI-23-7
C 38	15.8	79.0	302	10	A1269988	A1269988 qt62a10.x
C 39	15.8	79.0	305	10	A1434249	A1434249 t133d02.x
C 40	15.8	79.0	360	10	A1800661	A1800661 tc12f07.x
C 41	15.8	79.0	400	11	BF395655	BF395655 UI-R-CM0-
C 42	15.8	79.0	401	11	BF395115	BF395115 UI-R-CM0-
C 43	15.8	79.0	407	11	BF998712	BF998712 OVO-GN021
C 44	15.8	79.0	408	13	A2914287	A2914287 RPCI-24-1
C 45	15.8	79.0	414	11	BF816874	BF816874 MR2-C1012

ALIGNMENTS

RESULT 1
BG386035 353 bp mRNA EST 12-MAR-2001
LOCUS 602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5', mRNA sequence.
DEFINITION BG386035
ACCESSION BG386035
VERSION BG386035.1 GI:13279481
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 353)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1307 row: h column: 02
High quality sequence stop: 276.
Location/Qualifiers
1. 353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"

FEATURES
source

/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT
ORIGIN

80 a 111 c 122 g 40 t

Query Match 100.0%; Score 20; DB 11; Length 353;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attaccacccccctcgga 20

||||| ||||| ||||| ||||| |||||

Db 245 ATTACCACCCCTCGGGA 264

RESULT 2

AA311787

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 273)

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult

,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White

,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,

Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald

,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,

Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,

Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,

Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

Small,K.V., Spraggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,

Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.

, Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,

Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,

Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon

,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and

Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavetigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..273

/organism="Homo sapiens"

/db_xref="ATCC (inhost):158998"

/db_xref="taxon:9606"

/clone_lib="Jurkat T-cells V1"

/cell_type="T-lymphocyte"

/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT
ORIGIN

60 a 111 c 59 g 39 t 4 others

Query Match 90.0%; Score 18; DB 10; Length 273;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attaccacccccctcgga 20

||||| ||||| ||||| ||||| |||||

Db 129 ATTACCNNCCCTCGGGA 148

RESULT 3

AW588384/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 435)

McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.

, Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.

, Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagaris,I., R.

, Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe

,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.

, Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and

Wilson,R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

The Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Dr. John Hawdon of Yale University

(john.hawdon@yale.edu). DNA Sequencing by: Washington University

Genome Sequencing Center, St. Louis.

Seq primer: -40RP from Gibco

High quality sequence stop: 421.

Location/Qualifiers

1..435

/organism="Ancylostoma caninum"

/strain="Shanghai strain"

/db_xref="taxon:29170"

/clone_lib="Hawdon Ancylostoma caninum L3"

/sex="female and male"

/dev_stage="3rd stage larva (L3)"

/lab_host="E. coli, Xi-1 Blue MRF' (Stratagene)"

/note="Vector: Lambda Uni-ZAP XR Vector (Stratagene);

Site_1: EcoRI (5'); Site_2: XhoI (3'); Oligo (dT) primed

library from L3 worms isolated from dog. cDNA was

constructed and cloned unidirectionally into the vector

from the EcoRI to the XhoI site. The library went

through one round of amplification."

BASE COUNT 102 a 102 c 63 g 168 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 435;

Best Local Similarity 94.7%; Pred. No. 2.1e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ttaccacccccctcgga 20

Db 223 ACTACCTCCCTCCCTCGGA 242

RESULT 8

BE684213 542 bp mRNA EST 25-APR-2001

LOCUS 184527 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.

DEFINITION BE684213

ACCESSION BE684213.1 GI:10071837

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

1 (bases 1 to 542)

REFERENCE

AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and Keeler,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGCACAT
BACKWARD: GTTTCCTCAGTCACGCG
Plate: 90 row: E column: 16
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..542
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pcwv SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 126 a 128 c 116 g 172 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 542;
Best Local Similarity 90.0%; Pred. No. 3.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 attaccaccccccctcgga 20
||||||| ||||| |||||

Db 467 ATTACCACCTCCCTTGGGA 486

RESULT 9

AW155439/c

LOCUS

DEFINITION mgie0017017f Rice blast infection stage cDNA library Oryza sativa/Pycicularia oryzae mixed EST library

ACCESSION AW155439

VERSION AW155439.1 GI:6224308

KEYWORDS

SOURCE

ORGANISM

Oryza sativa/Pycicularia oryzae mixed EST library.
Oryza sativa/Pycicularia oryzae mixed EST library.


```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; mixed EST libraries.
1 (bases 1 to 880)
Pauvaree R., Choi W. and Dean, R.A.
Identification and characterization of genes expressed by the rice
blast pathogen and rice during infection stage
Unpublished (1999)
Contact: Dean, R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: T3 primer (AATTACCTCACTAAAGGG)
High quality sequence stop: 60.
FEATURES
source
1..880
Location/Qualifiers
/organism="Oryza sativa/Pyricularia oryzae mixed EST
library"
/strain="70-15"
/db_xref="taxon:105664"
/clone_lib="Rice blast infection stage cDNA library"
/dev_stage="Infection stage at 48 hour post-inoculation"
/note="Vector: pBlueScriptII SK(+); Vector: Rice
blast-infected leaves at 48 hour post-inoculation mRNA for
cDNA library construction."
BASE COUNT 213 a 239 c 202 g 225 t 1 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 880;
Best Local Similarity 90.0%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attaccacccccctcgga 20
|||||
Db 781 ATTTCCACCCCTCGAGA 762

RESULT 10
LOCUS BF106598/8 896 bp mRNA EST 19-OCT-2000
DEFINITION 601823238R1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4043264 3',
mRNA sequence.
ACCESSION BF106598
VERSION BF106598.1 GI:10889124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 896)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM865 row: k column: 09
High quality sequence stop: 4
High quality sequence stop: 5.
FEATURES
source
1..896
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4043264"

/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDMR-LIB (Clontech); Site_1:
SfiI (ggccattggcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 217 a 183 c 265 g 230 t 1 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 896;
Best Local Similarity 90.0%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attaccacccccctcgga 20
|||||
Db 470 ATTAACACCCCTCGGGA 451

RESULT 11
LOCUS BE216883
DEFINITION HV_CEB0011P02f Hordeum vulgare seedling green leaf EST library
HVCDNA0005 (Erysiphe infected & control) Hordeum vulgare cDNA clone;
HV_CEB0011P02f, mRNA sequence.
ACCESSION BE216883
VERSION BE216883.1 GI:8904495
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
Triticeae; Hordeum.
1 (bases 1 to 969)
Wing, R., Close, T.J., Klein, H., Wise, R., Begum, D., Frisch, D., Yu
Y., Anderson, H., Dale, J., Henry, D., Kern, S., Palmer, M., Rambo
T., Saski, C., Schwartz, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing, R.A.
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCTCACTAAAGGG
High quality sequence stop: 530.
FEATURES
source
1..969
Location/Qualifiers
/organism="Hordeum vulgare"
/cultivar="C116151 (M1a6)"
/db_xref="taxon:4513"
/clone="HV_CEB0011P02f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDNA0005 (Erysiphe infected & control)"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT 182 a 325 c 300 g 161 t 1 others
ORIGIN

```


Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
B., Swallier, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
MGI:377151

TITLE
JOURNAL
COMMENT

Seq primer: -40RP from Gibco.
Location/Qualifiers
1. .368

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:616327"
/clone_lib="Stratagene mouse embryonic carcinoma (#937317
)"
/tissue_type="carcinoma"
/dev_stage="embryonic"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: oligo dT. P19 cell
line. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 73 a 103 c 101 g 91 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 368;
Best Local Similarity 94.4%; Pred. No. 5.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 taccacacccctcggga 20
||||| |||||||
Db 282 TACCACACCCCTCGGGA 265

RESULT 15
AA954701
LOCUS
DEFINITION AA954701 448 bp mRNA EST 07-JUL-1998
O08406.s1 NCI-CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1572850 3',
mRNA sequence.
ACCESSION AA954701
VERSION AA954701.1 GI:3118396
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 754 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1. .448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1572850"
/clone_lib="NCI-CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGAATTCGGCGCGCAATATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 167 a 90 c 63 g 128 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 448;
Best Local Similarity 94.4%; Pred. No. 5.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attaccacccctcg 18
||||| |||||||
Db 355 ATTACCCACCCCTTGS 372

Search completed: February 20, 2002, 10:25:49
Job time: 9857 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 11:01:42 : Search time 1729.17 Seconds
(without alignments)
133.567 Million cell updates/sec

Title: US-09-904-420A-5
Perfect score: 14
Sequence: 1 cctgctccgaggg (4)

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo_hum.*
- 31: em.htgo_inv.*
- 32: em.htgo_rod.*
- 33: em.htg_hum.*
- 34: em.htg_inv.*
- 35: em.htg_rod.*
- 36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14	100.0	740	10	MUSNICB	L20476 Mus musculu
C 2	14	100.0	1115	9	HSTCF1D	Z47361 H.sapiens T
C 3	14	100.0	1165	9	HSTCF1E	Z47362 H.sapiens T
C 4	14	100.0	1250	9	BC006119	BC006119 Homo sapi
C 5	14	100.0	1254	9	HSTCF1A	X59869 Human TCF-1
C 6	14	100.0	1630	9	AK023286	AK023286 Homo sapi
C 7	14	100.0	2093	9	HUMMPF	D49441 Human mRNa
8	14	100.0	2114	9	HSU40434	U40434 Human mesot
9	14	100.0	2129	6	E07943	E07943 cDNA encodi
10	14	100.0	2129	6	I89985	I89985 Sequence 34
11	14	100.0	2138	6	AR100763	AR100763 Sequence
12	14	100.0	2138	6	AR119934	AR119934 Sequence
13	14	100.0	2162	9	BC009272	BC009272 Homo sapi
14	14	100.0	2433	9	BC003512	BC003512 Homo sapi
15	14	100.0	2589	4	SSR236925	AJ236925 Sus scrofa
C 16	14	100.0	2814	9	HSTCF1B	X59870 Human TCF-1
C 17	14	100.0	2855	9	HSTCF1G	X63901 Homo sapien
C 18	14	100.0	2910	9	HSTCF1C	X59871 Human TCF-1
19	14	100.0	4049	10	MM043298	U43298 Mus musculu
20	14	100.0	4078	10	BC008516	BC008516 Mus muscu
21	14	100.0	9356	4	AF242855	AF242855 Bos tauru
C 22	14	100.0	27645	9	HS56665	AL023577 Human DNA
C 23	14	100.0	37307	9	HS335H7	AL031258 Human DNA
C 24	14	100.0	48265	9	AC000388	AC000388 Genomic s
C 25	14	100.0	83909	9	HS032498	AL121942 Human DNA
26	14	100.0	84544	2	AC009012	AC009012 Homo sapi
27	14	100.0	86914	2	AC011336	AC011336 Homo sapi
28	14	100.0	99141	2	AP003704	AP003704 Oryza sat
29	14	100.0	110000	2	HSS171M_0	AJ239326 Homo sapi
C 30	14	100.0	118429	9	AL158017	AL158017 Homo sapi
C 31	14	100.0	138177	9	AL139286	AL139286 Human DNA
32	14	100.0	139226	2	AC084233	AC084233 Homo sapi
C 33	14	100.0	149816	2	AL359540	AL359540 Homo sapi
34	14	100.0	154616	2	AC067772	AC067772 Homo sapi
35	14	100.0	159420	2	AC009017	AC009017 Homo sapi
36	14	100.0	161571	2	AL359186	AL359186 Homo sapi
37	14	100.0	163427	9	AC009053	AC009053 Homo sapi
38	14	100.0	166042	2	AC068793	AC068793 Homo sapi
C 39	14	100.0	166946	2	AP001008	AP001008 Homo sapi
C 40	14	100.0	171520	2	AP001637	AP001637 Homo sapi
41	14	100.0	171825	2	AC011324	AC011324 Homo sapi
42	14	100.0	176218	2	AC009153	AC009153 Homo sapi
C 43	14	100.0	177387	2	AC090046	AC090046 Mus muscu
44	14	100.0	177580	2	AL590710	AL590710 Homo sapi
C 45	14	100.0	178266	2	AC068792	AC068792 Homo sapi

ALIGNMENTS

RESULT 1

MUSNICB	MUSNICB	740 bp	mRNA	ROD	08-JUN-1995
LOCUS	Mus musculus 125 kDa nicein (Nicb1) mRNA sequence.				
DEFINITION	L20476				
ACCESSION	L20476.1	GI:854646			
VERSION	laminin; nicein.				
KEYWORDS	Mus musculus strain C57BL/C	cdna to mRNA.			
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 740)				
AUTHORS	Aberdam,D., Aguzzi,A., Baudoin,C., Galliano,M.F., Mattei,M.G., Pisani-Spadafora,A., Ortonne,J.P. and Meneguzzi,G.				
TITLE	Assignment of mouse nicein genes to chromosomes 1 and 18				
JOURNAL	Mamm. Genome 5 (4), 229-233 (1994)				
MEDLINE	94281750				
REFERENCE	2 (bases 1 to 740)				
AUTHORS	Aberdam,D., Aguzzi,A., Baudoin,C., Galliano,M.F., Ortonne,J.P. and Meneguzzi,G.				

TITLE Developmental expression of nicein adhesion protein (laminin-5) subunits suggests multiple morphogenic roles

JOURNAL Cell Adhes. Commun. 2 (2), 115-129 (1994)

MEDLINE 94363405

REFERENCE 3 (bases 1 to 740)

AUTHORS Aberdeen, D.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-1994) Daniel Aberdeen, Faculte de Medicine, INSREM U395, Ave de Valombrese 06107, Nice Cedex 2, France

COMMENT On Jun 8, 1995 this sequence version replaced gi:833764.

FEATURES

source

1..740

/organism="Mus musculus"

/db_xref="taxon:10090"

<1..>740

/codon_start=1

/product="nicein"

/protein_id="AA068092.1"

/db_xref="GI:854647"

/translation="HMLRMTAFPPVACLPGLCAQACSRGACYPVPGDLLIGRTQL LRASSTGLTKPETYCTQYQWQMKCKDCSRLPRILGARKDMPCEETGRCLCLPNV VPKCDQATPHSHKLASGACWCPDPNRLSSQCNQFTGQCLVKGGLAALTCSAAI RQCPDQYGHVATGCRACDCDFRGTEGPKASGRCLCALAFTGPRCDQCQRGHCDR YPVCVACHSCFOAYDTLQSOARLHSLR"

BASE COUNT 139 a 227 c 206 g 168 t

ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 740;

Best Local Similarity 100.0%; Pred. No. 6e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

|||||

Db 71 CCTGCTCCCGAGG 84

RESULT 2

HSTCFD/c

LOCUS HSTCFD 1115 bp mRNA PRI 09-JAN-1995

DEFINITION H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.

ACCESSION 247361

VERSION 247361.1 GI:619881

KEYWORDS splice form D; T cell factor 1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1115)

AUTHORS Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.

TITLE The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon 1XA

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1115)

AUTHORS van de Wetering, M., Oosterwegel, M., Holsteg, F., Dooyes, D., Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.

TITLE The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization

JOURNAL J. Biol. Chem. 267 (12), 8530-8536 (1992)

MEDLINE 92235082

REFERENCE 3 (bases 1 to 1115)

AUTHORS Ballhausen, W.G.

TITLE Direct Submission

JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES

source

1..1115

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="lymphoma"

/cell_type="T-lymphocyte"

/cell_line="Jurkat"

1..1115

/gene="TCF-1"

/standard_name="T cell factor 1 splice form D"

/citation=[1]

/function="high mobility group box transcription factor"

/evidence=experimental

1..1115

/gene="TCF-1"

2..754

/standard_name="T cell factor 1 splice form D"

/citation=[1]

/codon_start=1

/product="T cell factor 1 splice form D"

/protein_id="CAA87439.1"

/db_xref="GI:619882"

/translation="MYKETVYSAFNLLMHYPPSGAGHQPPPLHKANOPPHGVPO LSLYGFHNSPHTPAPADISQVHRPLOTDLGPFYSLTSGSMGQLPHTVSWFTHPS LMLGSGVCPHPAAIPHPAIVPSPCKOELQDFDLNLTQAESKAEAKKPTIKKPLNA FMYLWKEKRAKVIACETLKESAAINQILGRWHALSREQAKYYELARKERLIQMQLY PWSARDNTGKKRRSRKHOESTIDNSLHYS"

731..751

/gene="TCF-1"

/product="alternative ORF specific for TCF-1D"

BASE COUNT 272 a 407 c 265 g 171 t

ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1115;

Best Local Similarity 100.0%; Pred. No. 5.6e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

|||||

Db 69 CCTGCTCCCGAGG 56

RESULT 3

HSTCFE/c

LOCUS HSTCFE 1165 bp mRNA PRI 09-JAN-1995

DEFINITION H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.

ACCESSION 247362

VERSION 247362.1 GI:619883

KEYWORDS splice form E; T cell factor 1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1165)

AUTHORS Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.

TITLE The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon 1XA

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1165)

AUTHORS van de Wetering, M., Oosterwegel, M., Holsteg, F., Dooyes, D., Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.

TITLE The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization

JOURNAL J. Biol. Chem. 267 (12), 8530-8536 (1992)

MEDLINE 92235082

REFERENCE 3 (bases 1 to 1165)

AUTHORS Ballhausen, W.G.

TITLE Direct Submission

JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES

source

1..1165

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="lymphoma"

```
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1165
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1165
/gene="TCF-1"
2. .1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKETVYSAFNLLMHPSPGAGQHPPQPPPLHKANQPPHGVPPQ
LSLYHFNPHPTAPADISQVHRPLQTPDLSGFSLTSGSMGQLPHTVSMFTHP
LMLSGVPGHPAAIPHPAIVPSPGKOELOPDRNLKTOAESKAEKAKPTLKPLNA
FMYLNKEMRAKVIAETLKESAAINQLGRWHALSREEQAKYELARKERQLHMOLY
PGWARDNGKRRRSREKHQESPTDPSGPKKCRAREGLAQOTDWCGRKCKJRY
LPGRCRCPVPSDDSLCCGSPAPQSPSYHLLPRPTELLTSPAERHLHPQVSP
LSASQPGPHRPPAAPACRAHRYSNRLNDRWPSRRHTPGRLQETPT"
803. .1141
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||
Db 69 cctgctcccgagg 56

RESULT 4
BC006119/c
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form A). 12-JUL-2001
DEFINITION BC006119 1250 bp mRNA PRI
ACCESSION BC006119
VERSION BC006119.1 GI:13543960
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1250)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NTH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgaps@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
```

```
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1165
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1165
/gene="TCF-1"
2. .1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKETVYSAFNLLMHPSPGAGQHPPQPPPLHKANQPPHGVPPQ
LSLYHFNPHPTAPADISQVHRPLQTPDLSGFSLTSGSMGQLPHTVSMFTHP
LMLSGVPGHPAAIPHPAIVPSPGKOELOPDRNLKTOAESKAEKAKPTLKPLNA
FMYLNKEMRAKVIAETLKESAAINQLGRWHALSREEQAKYELARKERQLHMOLY
PGWARDNGKRRRSREKHQESPTDPSGPKKCRAREGLAQOTDWCGRKCKJRY
LPGRCRCPVPSDDSLCCGSPAPQSPSYHLLPRPTELLTSPAERHLHPQVSP
LSASQPGPHRPPAAPACRAHRYSNRLNDRWPSRRHTPGRLQETPT"
803. .1141
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||
Db 69 cctgctcccgagg 56

RESULT 4
BC006119/c
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form A). 12-JUL-1991
DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form A).
ACCESSION X59869.X5327
VERSION X59869.1 GI:36785
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1254)
AUTHORS van de Wetering,M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
2 (bases 1 to 1254)
van de Wetering,M., Oosterwegel,M., Dooljes,D. and Clevers,H.
Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box
EMBO J. 10 (1), 123-132 (1991)
91114695
COMMENT See also X59869.X59871.
FEATURES
source
1. .1254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat"
/clone_lib="cdna"
/clone="fTCF-1a"
1. .1254
/gene="TCF-1"

mRNA
```

Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LJNL at: <http://image.llnl.gov>

Series: IRAL Plate: 18 Row: d Column: 3

This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.

Location/Qualifiers

1. .1250

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3505629"

/tissue_type="Placenta, choriocarcinoma"

/clone_lib="NIH_MGC_21"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

<1. .200

/codon_start=3

/product="Unknown (protein for IMAGE:3505629)"

/protein_id="AAH06119.1"

/db_xref="GI:13543961"

/translation="WERNRGQTPRTHLAKVLKRGDPFKCTGVVLGWPWEGRSNIHDDVY

PPACPPRAOSVPCVKVKGQLL"

BASE COUNT 318 a 381 c 285 g 266 t

ORIGIN

Query Match 100.0%; Score 14; DB 9; Length 1250;

Best Local Similarity 100.0%; Pred. No. 5.5e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

|||||

Db 287 cctgctcccgagg 274

RESULT 5

HSTCFJA/c

LOCUS Human TCF-1 mRNA 1254 bp mRNA PRI 14-JUN-1991

DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form A).

ACCESSION X59869.X5327

VERSION X59869.1 GI:36785

KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1254)

AUTHORS van de Wetering,M.

TITLE Direct Submission

JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

2 (bases 1 to 1254)

van de Wetering,M., Oosterwegel,M., Dooljes,D. and Clevers,H. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box

EMBO J. 10 (1), 123-132 (1991)

91114695

COMMENT See also X59869.X59871.

FEATURES

source

1. .1254

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="T lymphocyte"

/cell_line="Jurkat"

/clone_lib="cdna"

/clone="fTCF-1a"

1. .1254

/gene="TCF-1"

mRNA

us-09-904-420a-5.ige

Thu Feb 21 07:25:58 2002

```

/db_xref="taxon:9606"
/clone="OVARC1000008"
/clone_lib="OVARC1"
/tissue_type="ovary, tumor tissue"
/note="cloning vector: pME18SFL3"
529..1008
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14510.1"
/db_xref="GI:10435157"
/translat="MRNARSIRPARAESWRSLGCLFNQDALLSKPHCPAGRTCTPTAA
ASGPRGAGHSKDQGRSPKHVCHVHRKQCPAEFTDSRGVHNAASISSQVGPFGH
LVARAAHKREALVSPPTLQVNAQAKLRAPKAPRTPEALGRGPRAGEWKIGCF"
BASE COUNT      329 a  585 c  466 g   250 t
ORIGIN

Query Match      100.0%; Score 14; DB 9; Length 1630;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctccgagg 14
|||||
Db 497 CCTGCTCCGAGG 484

RESULT 7
HUMMPF 2093 bp mRNA PRI 01-FEB-2000
LOCUS Human mRNA for pre-pro-megakaryocyte potentiating factor, complete cds.
ACCESSION D49441
VERSION D49441.1 GI:1129078
KEYWORDS MPF; megakaryocyte potentiating factor.
SOURCE Homo sapiens pancreatic cancer cell cell_line:HPC-y5 cDNA to mRNA, clone:PKP027.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2093)
AUTHORS Kojima,T., Oh-eda,M., Hattori,K., Taniguchi,Y., Tamura,M., Ochi,N.
and Yamaguchi,N.
TITLE Molecular cloning and expression of megakaryocyte potentiating
factor cDNA
J Biol. Chem. 270 (37), 21984-21990 (1995)
J5594969
2 (bases 1 to 10)
AUTHORS Kojima,T.
JOURNAL Unpublished (1996)
REFERENCE 3 (bases 1 to 2093)
AUTHORS Kojima,T.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1995) to the DDBJ/EMBL/GenBank databases. Tetsuo
Kojima, Chugai Pharmaceutical Co.Ltd., Fuji Golemba Research
Laboratories, 1-135 Komakado, Golemba-shi, Shizuoka prefecture 412,
Japan (Tel:550-87-3411, Fax:550-87-5397)
FEATURES
Location/Qualifiers
1..2093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HPC-y5"
/cell_type="pancreatic cancer cell"
/clone="PKP027"
93..191
sig_peptide
CDS
93..191
/standard_name="MPF"
/citation=[2]
/codon_start=1
/product="pre-pro-megakaryocyte potentiating factor"
/protein_id="BAA08419.1"
/db_xref="GI:1374696"
/translat="MALPTAPRLPGSCGTPALCSLLFLFSLGNVQPSPTLAGETGOE
AAPLDGVLANPPNLSLSPPROLPGFCARFVSGIUSTERVRELAVALAQRNVKLSTQILR

/evidence=experimental
1..1254
/gene="TCF-1"
CDS 80..889
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1, splice form A"
/protein_id="CAA42526.1"
/db_xref="GI:36786"
/translat="MYKTVYSQFNLMLHYPPPSGAGHQPPQPPHLKANQPHGVQ
LSLGFHSFPHPTAPADISQVHRPLQPDLSGFSLTSGSMGLPHTVSWFTHPS
LMLGFCVCPHPAAIPHPAIVPPSGQELQPDFRLNLTQAESKAEEKAKPTIKPLNA
FMYLWKKRAKVIACETLKESAAIINILGRWHALSREOAKYELARKERQHLMOLY
PWSARDNYGKKRKRREKHQESTETINWPELKDGNQESLSMSSSSSPA"
539..769
/gene="TCF-1"
/note="HMG box"
810
/gene="TCF-1"
/note="alternative splice site"
BASE COUNT      320 a  391 c  335 g   208 t
ORIGIN

Query Match      100.0%; Score 14; DB 9; Length 1254;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctccgagg 14
|||||
Db 147 CCTGCTCCGAGG 134

RESULT 6
AK023286/c
LOCUS Homo sapiens CDNA FLJ13224 fis, clone OVARC1000008.
DEFINITION AK023286
ACCESSION AK023286.1 GI:10435156
VERSION AK023286.1 GI:10435156
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_lib:OVARC1
clone:OVARC1000008.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wadatsuma,M., Hosoiri,F., Kaku,F., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 1630)
Isodai,T. and Otsuki,T.
Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..1630
/organism="Homo sapiens"

```


CLAHRLSEPPEDLDALPLDLLLFLNPDPAFSGPOACTRFFSRITKANVOLLPRGAPERQ RLPLAALACWVGRLSLEADVRALGLGACDLPGRFAESAENVLLPRLVSCPGPLDQD QQEAAARALQGGPPYPGPTWSVTMDALRGLLPVLGQPI IRSIQGVIAVWRORS RQPSWQPERTILRPRFRREVEKTCPSGKKAREIDESLIFYKKWSEACVDAALLAT QMDRVNAPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSL ETLKALLEVKNKGHMSPOVATLIDRFVKRGOLDKDTLDTLTAFFPGYLCSLSPLELS SVPPSIWVRPQDLTCDPRLDVLVPKARLAFONNMNGSEYFVKIQSFLLGAPTEDL KALSOQVNSMDLATPMKLRDVAFLPVAEVOKLLGPHVEGLKAEERHVRPVDWILRQ RQDLDTLGLGGGPGNGYLVLDSVQEALSGTPLLGGPVLTVLALLLASTLA" 192. .950 /note="33kDa-MPF" /product="megakaryocyte potentiating factor"		sig_peptide mat_peptide misc_feature misc_feature misc_feature misc_feature polyA_signal polyA_site BASE COUNT ORIGIN		SPEELSSVPPSSINAVRPQDLTCDPRLDVLVPKARLAFONNMNGSEYFVKIQSFLLG APTEDLKALSOQVNSMDLATPMKLRDVAFLPVAEVOKLLGPHVEGLKAEERHVR DWILRQDLDTLGLGGGPGNGYLVLDSVQETUSGTPLLGGPVLTVLALLLA STLA"		100. .192 193. .1857 /product="mesothelin" 193. .978 /note="encodes putative cleaved N-terminal portion of mesothelin" 973. .978 /note="encodes putative proteolytic site" 979. .1983 /note="encodes putative membrane bound portion of mesothelin, bearing epitope(s) recognized by Mab K1" 1858. .1983 /note="encodes putative hydrophobic region for GPI anchorage" 2087. .2093 /note="putative" 2114 361 a 722 c 661 g 370 t	
Query Match 100.0%; Score 14; DB 9; Length 2093; Best Local Similarity 100.0%; Pred. No. 5.2e+02; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Query Match 100.0%; Score 14; DB 9; Length 2093; Best Local Similarity 100.0%; Pred. No. 5.2e+02; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Query Match 100.0%; Score 14; DB 9; Length 2114; Best Local Similarity 100.0%; Pred. No. 5.2e+02; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY 1 cctgctcccgagg 14 		QY 1 cctgctcccgagg 14 		QY 1 cctgctcccgagg 14 			
DB 542 CCTGCTCCCGAGG 555		DB 543 CCTGCTCCCGAGG 556		DB 543 CCTGCTCCCGAGG 556			
RESULT 8		RESULT 9		RESULT 9			
HSU040434 2114 bp mRNA PRI 19-JAN-1996		E07943 2129 bp RNA PAT 29-SEP-1997		E07943 2129 bp RNA PAT 29-SEP-1997			
LOCUS Human mesothelin or CAK1 antigen precursor mRNA, complete cds.		LOCUS cDNA encoding a polypeptide having megakaryocyte potentiating activity.		LOCUS cDNA encoding a polypeptide having megakaryocyte potentiating activity.			
ACCESSION U040434		ACCESSION E07943		ACCESSION E07943			
VERSION U040434.1 GI:1145723		VERSION E07943.1 GI:2176075		VERSION E07943.1 GI:2176075			
KEYWORDS 96133892		KEYWORDS JP 1994225767-A/1.		KEYWORDS JP 1994225767-A/1.			
SOURCE human.		SOURCE Homo sapiens.		SOURCE Homo sapiens.			
ORGANISM Homo sapiens		ORGANISM Homo sapiens		ORGANISM Homo sapiens			
REFERENCE 1 (bases 1 to 2114)		REFERENCE 1 (bases 1 to 2129)		REFERENCE 1 (bases 1 to 2129)			
AUTHORS Chang,K. and Pastan,I.		AUTHORS Yamaguchi,M., Kojima,T., Oeda,M. and Hattori,A.		AUTHORS Yamaguchi,M., Kojima,T., Oeda,M. and Hattori,A.			
TITLE Molecular cloning of mesothelin, a differentiation antigen present on mesothelium, mesotheliomas, and ovarian cancers		TITLE GENE CODING MAGKARYOCYTE AMPLIFIER		TITLE GENE CODING MAGKARYOCYTE AMPLIFIER			
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (1), 136-140 (1996)		JOURNAL Patent: JP 1994225767-A 1 16-AUG-1994;		JOURNAL Patent: JP 1994225767-A 1 16-AUG-1994;			
REFERENCE 2 (bases 1 to 2114)		REFERENCE 2 (bases 1 to 2114)		REFERENCE 2 (bases 1 to 2114)			
AUTHORS Chang,K.		AUTHORS Chang,K.		AUTHORS Chang,K.			
TITLE Direct Submission		TITLE Submitted (09-NOV-1995) Kai Chang, Laboratory of Molecular Biology,		TITLE Submitted (09-NOV-1995) Kai Chang, Laboratory of Molecular Biology,			
JOURNAL National Cancer Institute, Building 37, Room 4B19, 37 Convent		JOURNAL National Cancer Institute, Building 37, Room 4B19, 37 Convent		JOURNAL National Cancer Institute, Building 37, Room 4B19, 37 Convent			
FEATURES source		FEATURES source		FEATURES source			
1. .2114		1. .2114		1. .2129			
/organism="Homo sapiens"		/organism="Homo sapiens"		/organism="Homo sapiens"			
/db_xref="taxon:9606"		/db_xref="taxon:9606"		/db_xref="taxon:9606"			
/clone="pcD3CAK1-9"		/clone="pcD3CAK1-9"		/clone="pcD3CAK1-9"			
/chromosome="16"		/chromosome="16"		/chromosome="16"			
/cell_line="KB, HeLa, OVCAR-3"		/cell_line="KB, HeLa, OVCAR-3"		/cell_line="KB, HeLa, OVCAR-3"			
/cell_type="HeLa"		/cell_type="HeLa"		/cell_type="HeLa"			
/tissue_type="ovarian cancers and squamous cell carcinoma."		/tissue_type="ovarian cancers and squamous cell carcinoma."		/tissue_type="ovarian cancers and squamous cell carcinoma."			
100. .1986		100. .1986		100. .1986			
/codon_start=1		/codon_start=1		/codon_start=1			
/product="mesothelin or CAK1 antigen precursor"		/product="mesothelin or CAK1 antigen precursor"		/product="mesothelin or CAK1 antigen precursor"			
/protein_id="AAC50348.1"		/protein_id="AAC50348.1"		/protein_id="AAC50348.1"			
/db_xref="GI:1145724"		/db_xref="GI:1145724"		/db_xref="GI:1145724"			
/translation="MALQRLDPCWSCDRGSLLELFSIGWVHARTLAGETGESA PLGGVTTTPHNISSLPRLQGFPAEVSGLSTERVRELAVALAKNKKLSTEQRLCL AHRUSEPDEDLALPDLILLFNPDAFSGPOACTRFFSRITKANVOLLPRGAPERQRL LPAALACWVGRLSLEADVRALGLGACDLPGRFAESAENVLLPRLVSCPGPLDQDQ EAAARALQGGPPYPGPTWSVTMDALRGLLPVLGQPI IRSIQGVIAVWRORSRD RQPSWQPERTILRPRFRREVEKTCPSGKKAREIDESLIFYKKWSEACVDAALLATQ MDRVNAPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE TDRVNAIPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE LTKALLEVDKGMSPQAPRRPLQVATLIDRFVKRGOLDKDTLDTLTAFFPGYLCSL		/translation="MALQRLDPCWSCDRGSLLELFSIGWVHARTLAGETGESA PLGGVTTTPHNISSLPRLQGFPAEVSGLSTERVRELAVALAKNKKLSTEQRLCL AHRUSEPDEDLALPDLILLFNPDAFSGPOACTRFFSRITKANVOLLPRGAPERQRL LPAALACWVGRLSLEADVRALGLGACDLPGRFAESAENVLLPRLVSCPGPLDQDQ EAAARALQGGPPYPGPTWSVTMDALRGLLPVLGQPI IRSIQGVIAVWRORSRD RQPSWQPERTILRPRFRREVEKTCPSGKKAREIDESLIFYKKWSEACVDAALLATQ MDRVNAPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE TDRVNAIPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE LTKALLEVDKGMSPQAPRRPLQVATLIDRFVKRGOLDKDTLDTLTAFFPGYLCSL		/translation="MALQRLDPCWSCDRGSLLELFSIGWVHARTLAGETGESA PLGGVTTTPHNISSLPRLQGFPAEVSGLSTERVRELAVALAKNKKLSTEQRLCL AHRUSEPDEDLALPDLILLFNPDAFSGPOACTRFFSRITKANVOLLPRGAPERQRL LPAALACWVGRLSLEADVRALGLGACDLPGRFAESAENVLLPRLVSCPGPLDQDQ EAAARALQGGPPYPGPTWSVTMDALRGLLPVLGQPI IRSIQGVIAVWRORSRD RQPSWQPERTILRPRFRREVEKTCPSGKKAREIDESLIFYKKWSEACVDAALLATQ MDRVNAPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE TDRVNAIPFTYEQDLVKHLKDELQPGYPESVIOHLGYFLFKMSPEEDIRKNWVTSLE LTKALLEVDKGMSPQAPRRPLQVATLIDRFVKRGOLDKDTLDTLTAFFPGYLCSL			
CDS		CDS		CDS			

FT /clone='pkPO27'
FT 1..96
FT 97..1965
FT /product='polypeptide having megakaryocyte FT
FT potentiating
FT activity'
FT 1966..2129
FT 3'UTR
FT misc_feature
FT 1873
FT /note='The nucleotide at position 1873 of cDNA
FT in other
FT clone pkPO21 is A'.
FEATURES
source Location/Qualifiers
1..2129
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 389 a 716 c 657 g 367 t
ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 2129;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctcccgagg 14
|||||
Db 546 CCTGCTCCGAGG 559

RESULT 10
189985
LOCUS 189985 2129 bp DNA PAT 10-AUG-1998
DEFINITION Sequence 34 from patent US 5723318.
ACCESSION I89985
VERSION I89985.1 GI:3409925
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2129)
AUTHORS Yanaguchi,N., Kojima,T., Oh-eda,M. and Hattori,K.
TITLE DNA coding for megakaryocyte potentiator
JOURNAL Patent: US 5723318-A 34 03-MAR-1998;
FEATURES
source Location/Qualifiers
1..2129
/organism='unknown'
BASE COUNT 389 a 716 c 657 g 367 t
ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 2129;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctcccgagg 14
|||||
Db 546 CCTGCTCCGAGG 559

RESULT 11
AR100763
LOCUS AR100763 2138 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083502.
ACCESSION AR100763
VERSION AR100763.1 GI:12811561
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2138)
AUTHORS Pastan,I. and Chang,K.
TITLE Mesothelium antigen and methods and kits for targeting it
JOURNAL Patent: US 6083502-A 1 04-JUL-2000;
FEATURES
source Location/Qualifiers

source 1..2138
/organism='unknown'
BASE COUNT 385 a 722 c 661 g 370 t
ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 2138;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctcccgagg 14
|||||
Db 543 CCTGCTCCGAGG 556

RESULT 12
AR119934
LOCUS AR119934 2138 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6153430.
ACCESSION AR119934
VERSION AR119934.1 GI:14102633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2138)
AUTHORS Pastan,I. and Chang,K.
TITLE Nucleic acid encoding mesothelin, a differentiation antigen present on mesothelium, mesotheliomas and ovarian cancers
JOURNAL Patent: US 6153430-A 1 28-NOV-2000;
FEATURES
source Location/Qualifiers
1..2138
/organism='unknown'
BASE COUNT 385 a 722 c 661 g 370 t
ORIGIN

Query Match 100.0%; Score 14; DB 6; Length 2138;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctactcccgagg 14
|||||
Db 543 CCTGCTCCGAGG 556

RESULT 13
BC009272
LOCUS BC009272 2162 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:10273 IMAGE:3957372, mRNA, complete cds.
ACCESSION BC009272
VERSION BC009272.1 GI:14424504
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2162)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (I.M.L.)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

COMMENT

Contact: MGC help desk
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Nees, Pawan Pandoh, Anna-Liisa Prabhhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasia van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL plate: 13 Row: a Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 1129078.

FEATURES
source

Location/Qualifiers
1..2433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:10686 IMAGE:3611296"
/tissue_type="Pancreas, adenocarcinoma"
/clone_lib="NIH_MGC_39"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
418..2283
/codon_start=1
/product="mesothelin"
/protein_id="AAH03512.1"
/db_xref="GI:13097582"
/translation="MALPTARPLGSGTALGSLFLFSLGWVOPSRRLAGETGOA
AAPLDGLVLANPNISLSPLROLGFCFAEVSGLSTERVRELAVALAKNVKLSLEQLRC
LAHRLSEPPEDIDALPLDLLFLNPDAFSGPQACTHFRSRTKANVDLLPGAPRQR
LILPALACWVGRLSLLSEADVRALGLGLACDIPGRFVAESAEVLLPLRVSCPGIPDQD
QEAARAALQGGPPYPGSTWSTMDALRGILFVIGQPIIKRSIPQGIWAARQSSR
DPSWRQPERTILRRFRREVEKTACPSGKKAIEDLSLIFYKKWIEACVDAALLATQ
MDRVNAIPFTYEQLDVLKHLDELYPQGPESVIOHLGYLFRKMSPEDIRKNVNTSL
TLKALLVKNKGHEMSPQVATLIDRFVKGRLQDKDTLDTITAFPGYLCLSLSPLELS
VPPSSIAWRQDDLTCDPRQDLVLYPKARLAFONMNGSEYFVKIQSFLGGAPTEDLK
ALUSQNVSMDLATFMKLTDAVPLTVAEVOKLIGPHVEGLKAEERHRPVRDWILRQ
ODDLDTLGLGLOGGIPNGTIVLDLSVOEALSGTFLGPGPVLTVLALLASTLA"

CDS

BASE COUNT 431 a 823 c 755 g 424 t
ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 2433;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cctgctccccgagg 14
DB 864 CCTGCTCCCGAGG 877

RESULT 15
LOCUS

SSR236925 2589 bp mRNA MAM 18-AUG-2000
Sus scrofa mRNA for hypothetical protein (5'; clone ID8).
AJ236925
AJ236925.1 GI:4186137
KEYWORDS
SOURCE
ORGANISM
REFERENCE
pig.
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 2589)

Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hghri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, G.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 15 Row: j Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7108357.

FEATURES
source

Location/Qualifiers
1..2162
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:10273 IMAGE:3957372"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
117..1985
/codon_start=1
/product="Unknown (protein for MGC:10273)"
/protein_id="AAH09272.1"
/db_xref="GI:14424505"
/translation="MALPTARPLGSGTALGSLFLFSLGWVOPSRRLAGETGOE
AAPLDGLVLANPNISLSPLROLGFCFAEVSGLSTERVRELAVALAKNVKLSLEQLR
LAHRLSEPPEDIDALPLDLLFLNPDAFSGPQACTHFRSRTKANVDLLPGAPRQR
LILPALACWVGRLSLLSEADVRALGLGLACDIPGRFVAESAEVLLPLRVSCPGIPDQD
QEAARAALQGGPPYPGSTWSTMDALRGILFVIGQPIIKRSIPQGIWAARQSSR
DPSWRQPERTILRRFRREVEKTACPSGKKAIEDLSLIFYKKWIEACVDAALLAT
MDRVNAIPFTYEQLDVLKHLDELYPQGPESVIOHLGYLFRKMSPEDIRKNVNTSL
TLKALLVKNKGHEMSPQVATLIDRFVKGRLQDKDTLDTITAFPGYLCLSLSPLELS
VPPSSIAWRQDDLTCDPRQDLVLYPKARLAFONMNGSEYFVKIQSFLGGAPTEDLK
ALUSQNVSMDLATFMKLTDAVPLTVAEVOKLIGPHVEGLKAEERHRPVRDWILRQ
ODDLDTLGLGLOGGIPNGTIVLDLSVOEALSGTFLGPGPVLTVLALLASTLA"

CDS

BASE COUNT 416 a 721 c 659 g 366 t
ORIGIN
Query Match 100.0%; Score 14; DB 9; Length 2162;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cctgctccccgagg 14
DB 566 CCTGCTCCCGAGG 579

RESULT 14

BC003512 2433 bp mRNA PRI 12-JUL-2001
Homo sapiens, mesothelin, clone MGC:10686 IMAGE:3611296, mRNA,
complete cds.
ACCESSION BC003512
VERSION BC003512.1 GI:13097581
KEYWORDS
SOURCE
ORGANISM
REFERENCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2433)
AUTHORS
TITLE
JOURNAL
Submitted (20-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

REMARK

AUTHORS Cirera,S., Wintero,A.K. and Fredholm,M.
TITLE Why do we still find anonymous ESTs?
JOURNAL Mamm. Genome 11 (8), 689-693 (2000)
MEDLINE 20380830
REFERENCE 2 (bases 1 to 2589)
AUTHORS Cirera,S.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Cirera S., The Royal Veterinary and
Agricultural University, Department of Animal Science and Animal
Health, Division of Animal Genetics, Groennegaardsvej 3, 1870
Frederiksberg C, DENMARK

FEATURES
source Location/Qualifiers
1..2589
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="small intestine"
/clone_lib="directionally cloned cDNA in XLI-blue MRF"
/clone="ID8"
1..2589
/product="hypothetical protein"
BASE COUNT 715 a 642 c 608 g 605 t 19 others
ORIGIN

mRNA

Query Match 100.0%; Score 14; DB 4; Length 2589;
Best local Similarity 100.0%; Pred. No. 5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14
|||||
Db 2362 CCTGCTCCCGAGGG 2349

Search completed: February 20, 2002, 11:01:45
Job time: 12013 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:31:56 ; Search time 363.89 Seconds
(without alignments)
32.984 Million cell updates/sec

Title: US-09-904-420A-5
Perfect score: 14
Sequence: 1 cctgctcccgagg 14

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	14	100.0	724	22 AAH16880	Human cDNA clone (
C 2	14	100.0	1630	22 AAH17934	Human cDNA sequenc
C 3	14	100.0	2129	15 AAQ63972	Megakaryocyte pote
C 4	14	100.0	2138	18 AAQ91079	Human Cdk1 antigen
C 5	13	92.9	347	14 AAQ60393	Human brain expres
C 6	13	92.9	650	22 AAF93992	Primer specific fo
C 7	13	92.9	994	22 AAH94476	Human foetal cDNA,
C 8	13	92.9	1147	20 AAZ33452	Human prostate can
C 9	13	92.9	1383	20 AAX91681	Porphorymonas ging
C 10	13	92.9	1386	20 AAX91557	Porphorymonas ging
C 11	13	92.9	1668	20 AAV80620	Kidney injury asso

12	13	92.9	1830	20 AAV72023	Adenovirus pig4kN
13	13	92.9	1954	22 AAF60009	Human synthetase #
14	13	92.9	1990	21 AAZ36247	cDNA encoding a bo
15	13	92.9	2253	20 AAV72024	Adenovirus PNSFig4
C 16	13	92.9	2482	22 AAH14228	Human cDNA sequenc
C 17	13	92.9	3076	20 AAH02987	Human IL-1ra BAC c
C 18	13	92.9	3260	21 AAA09356	p75-NTR (neurotrop
C 19	13	92.9	3988	21 AAC68342	Rat insulin recept
C 20	13	92.9	4998	9 AAH81114	Non-A, non-B hepat
C 21	13	92.9	7263	22 AAH58179	Human polynucleoti
C 22	12.4	88.6	25	21 AAZ51453	Human NK2R promote
C 23	12.4	88.6	215	21 AAC09764	Human secreted pro
C 24	12.4	88.6	243	21 AAF21793	Human breast and o
C 25	12.4	88.6	247	14 AAQ61225	Human brain Expres
C 26	12.4	88.6	277	21 AAZ46505	Human exonuxlease
C 27	12.4	88.6	320	20 AAV87844	EST clone PA139.
C 28	12.4	88.6	320	22 AAH122161	Probe #12094 for g
C 29	12.4	88.6	320	22 AAH147459	Probe #16145 used
C 30	12.4	88.6	320	22 AAH07863	Probe #7854 used t
C 31	12.4	88.6	354	21 AAC03371	Human secreted pro
C 32	12.4	88.6	413	21 AAH87116	Rat hepatocyte car
C 33	12.4	88.6	422	14 AAQ59462	Human brain Expres
C 34	12.4	88.6	435	21 AAH42315	Human secreted exp
C 35	12.4	88.6	449	21 AAC98257	Human colon cancer
C 36	12.4	88.6	466	21 AAC00888	Human secreted pro
C 37	12.4	88.6	468	21 AAH99394	Arabidopsis thalia
C 38	12.4	88.6	486	22 AAH44622	Human protein enco
C 39	12.4	88.6	488	21 AAZ35041	Mouse CNRFB-1 rela
C 40	12.4	88.6	500	21 AAH65477	Porcine BAC-PiGF2-
C 41	12.4	88.6	509	21 AAC01713	Human secreted pro
C 42	12.4	88.6	516	19 AAV29354	Calcium ion channe
C 43	12.4	88.6	532	22 AAH97957	Murine 7-transmemb
C 44	12.4	88.6	538	22 AAH97961	Murine 7-transmemb
C 45	12.4	88.6	546	22 AAH97956	Murine 7-transmemb

ALIGNMENTS

RESULT 1
AAH16880/C
ID AAH16880 standard; cDNA; 724 BP.
XX AC AAH16880;
DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:16157.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-JUN-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

Thu Feb 21 07:25:58 2002

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1: SEQ ID 16157; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 724 BP; 134 A; 272 C; 219 G; 94 T; 5 other;

Query Match 100.0%; Score 14; DB 22; Length 724;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14
Db 497 CCTGCTCCCGAGG 484

RESULT 2
AAH17934/C
ID AAH17934 standard; cDNA; 1630 BP.
XX
AC AAH17934;
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:17699.
XX
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.

XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17699; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1630 BP; 329 A; 585 C; 466 G; 250 T; 0 other;

Query Match 100.0%; Score 14; DB 22; Length 1630;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14
Db 497 CCTGCTCCCGAGG 484

RESULT 3
AAQ63972
ID AAQ63972 standard; cDNA; 2129 BP.
XX
AC AAQ63972;
XX
XX 08-DEC-1994 (first entry)
XX Megakaryocyte potentiator (pKp027).
XX Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;
KW platelet; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 97..1965
FT /*tag= a
FT /product= Meg-Pot
FT 211..1962
FT /*tag= b
FT /note= "claim 1; 584 amino acid product"
FT 211..954
FT /*tag= c
FT /note= "claim 6; 248 amino acid product"
XX
XX WO9410312-A.
XX
XX 11-MAY-1994.
XX

```

PF 25-OCT-1993; 93WO-JP01540.
XX
XX 23-OCT-1992; 92JP-0286153.
PR 11-NOV-1992; 92JP-0301387.
PR 09-DEC-1992; 92JP-0329546.
XX
XX (CHUS ) CHUGAI SEIYAKU KK.
XX
XX Hattori K, Kojima T, Oh-eda M, Yamaguchi N;
XX
XX WPI: 1994-167467/20.
DR P-PSDB; AAR53992.
XX
XX New mega-karyocyte potentiator - for potential treatment of
XX thrombocytopenia
XX
XX Disclosure; Page 52-57; 74pp; Japanese.
XX
XX DNA encoding the 584 or 248 amino acid prod. has potential
XX use in treatment of thrombocytopenia and low platelet function.
XX
XX Sequence 2129 BP; 389 A; 715 C; 658 G; 367 T; 0 other;
SQ
Query Match 100.0%; Score 14; DB 15; Length 2129;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 cctgctcccgagg 14
Db 546 cctgctcccgagg 559
|||||
RESULT 4
AAT91079
ID AAT91079 standard; cDNA; 2138 BP.
XX
XX AAT91079;
AC
XX
XX 11-MAR-1998 (first entry)
DT
XX
XX Human CAK1 antigen (mesothelin) cDNA.
DE
XX
XX CAK1 antigen; mesothelin; tumour specific antigen; mesothelioma;
KW ovarian cancer; squamous cell cancer; gene therapy; diagnosis; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 100..1986
XX
XX sig_peptide 145..195
XX
XX /note= "putative signal sequence for membrane
XX insertion"
XX
XX polyA_signal 2087..2092
XX /tag= C
XX /note= "variant polyadenylation signal"
XX
XX W09725068-A2.
PN
XX
XX 17-JUL-1997.
PD
XX
XX 03-JAN-1997; 97WO-US00224.
XX
XX 05-JAN-1996; 96US-0010166.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chang K, Pastan I;
XX
XX WPI: 1997-372620/34.
XX P-PSDB; AAW26674.
DR
XX
XX Agents for targeting mesothelin, a tumour cell antigen - used for
XX the detection or inhibition of growth of e.g. mesotheliomas, ovarian
XX cancers and squamous cell carcinomas
XX
XX Claim 16; Page 58-60; 72pp; English.
XX
XX This cDNA clone codes for CAK1 antigen (see AAW26674) which is
XX found on mesothelium, mesotheliomas, ovarian cancers and some
XX squamous cell carcinomas. The antigen has been designated
XX mesothelin. The clone was isolated from a HeLa S3 cDNA library by
XX screening with monoclonal antibody K1, previously shown to react
XX with an antigen on human ovarian tumour OVCAR-3 cells. A 40 kDa
XX form (K1) of mesothelin appears to be derived from the 69 kDa
XX precursor by several processing steps. A claimed method for
XX specifically delivering an effector molecule to a tumour cell
XX bearing an antigen comprising at least 10 contiguous amino acids
XX of mesothelin involves: (a) providing a chimeric molecule comprising
XX the effector molecule attached to a targeting molecule that
XX specifically binds to mesothelin; and (b) contacting the tumour with
XX the chimeric molecule such that the chimeric molecule specifically
XX binds to a tumour cell. Also claimed is a method for inhibiting
XX mesothelin expression or activity by contacting mesothelin bearing
XX cells with inhibitory nucleic acids for the mesothelin gene. The
XX methods can be used to detect tumour cells and to inhibit the
XX growth of cells bearing mesothelin. Mesothelin-derived antigens
XX may be used in vaccines for the inhibition or prevention of
XX mesotheliomas or ovarian tumours. DNA encoding the mesothelin
XX antigen can be transfected into a mammal containing the mesothelin
XX screen for drugs useful in cancer treatment. Antisense
XX oligonucleotides can be used to inhibit mesothelin expression.
XX (All claimed).
XX
XX Sequence 2138 BP; 385 A; 722 C; 661 G; 370 T; 0 other;
SQ
Query Match 100.0%; Score 14; DB 18; Length 2138;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 cctgctcccgagg 14
Db 543 cctgctcccgagg 556
|||||
RESULT 5
AAQ60393/C
ID AAQ60393 standard; cDNA; 347 BP.
XX
XX AAQ60393;
AC
XX
XX 16-MAR-1994 (first entry)
DT
XX
XX Human brain Expressed Sequence Tag EST02392.
DE
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
XX transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX Homo sapiens.
OS
XX
XX W09316178-A.
PN
XX
XX 19-AUG-1993.
PD
XX
XX 12-FEB-1993; 93WO-US01294.
XX
XX 12-FEB-1992; 92US-0837195.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Adams MD, Moreno RF, Venter CJ;
XX
XX WPI: 1993-272882/34.
XX
XX

```

XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
PS Example 4; Page 327; 500pp; English.
XX
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST02392 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AA059041-Q61440.
XX
SQ Sequence 347 BP; 76 A; 87 C; 79 G; 101 T; 4 other;

Query Match 92.9%; Score 13; DA 14; Length 347;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ctgtctccgagg 14
|||||
DB 289 CTGCTCCGAGG 277

RESULT 6
AAF93992/C
ID AAF93992 standard; DNA; 650 BP.
XX
AC AAF93992;
XX
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 426.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS Synthetic.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 4; SEQ ID 426; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAF88817 - AAF88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the

CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 650 BP; 116 A; 200 C; 154 G; 170 T; 10 other;

Query Match 92.9%; Score 13; DA 22; Length 650;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctgtctccgagg 13
|||||
DB 30 CCTGCTCCGAGG 18

RESULT 7
AAH94476
ID AAH94476 standard; cDNA; 994 BP.
XX
AC AAH94476;
XX
DT 05-OCT-2001 (first entry)
XX
DE Human foetal cDNA, SEQ ID NO: 1163.
XX
KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW neutropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200155339-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02723.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tanq YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPI; 2001-465571/50.
DR P-PSDB; AAM06801.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Example 3; Page 656-657; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by

CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a full length cDNA which was assembled using expressed
CC sequence tags (ESTs) found to be expressed in human foetal tissue
CC cDNA libraries as seeds.
XX
SQ Sequence 994 BP; 320 A; 205 C; 176 G; 293 T; 0 other;

Query Match 92.9%; Score 13; DB 22; Length 994;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 13
|||||
Db 534 cctgctcccgagg 546
|||||

RESULT 8
AAZ33452
ID AAZ33452 standard; cDNA; 1147 BP.
XX
AC AAZ33452;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated EST 30.
XX
KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;
KW gene therapy; tissue specificity human; ss.
XX
OS Homo sapiens.
XX
PN DEJ9811193-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DF-1011193.
XX
PR 10-MAR-1998; 98DF-1011193.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
WPI: 1999-519628/44.
DR P-PSDB; AAY48249.
XX
PT New nucleic acid expressed at high level in prostatic tumor tissue and
PT encoded polypeptides, useful for treating cancer and screening for
PT therapeutic agents -
XX
PS Claim 3; 92; 166pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in prostatic tumor tissue and encode gene
CC products or their fragments. The products of the invention have
CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
CC identifying agents for treatment of prostatic cancer and (ii) for
CC therapy of prostate cancer, optionally where expressed by gene therapy
CC methods. (A) is also used to isolate full-length genes (for gene therapy)
CC and for recombinant production of (I), which can be used to raise
CC specific antibodies. (A) are identified by assembly of ESTs (expressed
CC sequence tags) before they are analyzed for expression pattern (tissue
CC specificity). This approach eliminates many of the false results, as
CC regards tissue specificity, associated with known methods that use
CC single (usually short) ESTs. AAZ33423-733476 represent expressed
CC sequence tags described in the method of the invention.
XX
SQ Sequence 1147 BP; 195 A; 424 C; 335 G; 193 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 1147;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 13
|||||
Db 1094 cctgctcccgagg 1106
|||||

RESULT 9
AAZ91681/c
ID AAZ91681 standard; DNA; 1383 BP.
XX
AC AAZ91681;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein PGI19 encoding DNA.
XX
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
XX
OS Porphyromonas gingivalis.
XX
PN WO9929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr TG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
WPI: 1999-385613/32.
DR P-PSDB; AAY34463.
XX
PT Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 12; Page 194; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 1383 BP; 337 A; 352 C; 424 G; 270 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 1383;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctccgagg 13
 ID AAX91557/c
 Db 169 CCTGCTCCGAGG 157

RESULT 10
 AAX91557/c
 ID AAX91557 standard; DNA; 1386 BP.

XX AAX91557;
 AC AAX91557;
 DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PGI19 ORF encoding DNA.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic; ds.

XX Porphyromonas gingivalis.

OS Porphyromonas gingivalis.

XX W09929870-Al.

XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

XX 30-JAN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

XX 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

XX 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

XX 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

XX WPI: 1999-385613/32.

DR P-PSDB; AAY34339.

XX Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis

PS Claim 12; Page 109; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic.
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

XX Sequence 1386 BP; 338 A; 352 C; 426 G; 270 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 1386;

Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctccgagg 13

Db 172 CCTGCTCCGAGG 160

RESULT 11

AAV80620/c
 ID AAV80620 standard; cDNA; 1668 BP.
 XX AAV80620;
 AC AAV80620;
 XX 01-MAR-1999 (first entry)
 DT 01-MAR-1999 (first entry)
 DE Kidney injury associated molecule HW076 cDNA clone.
 XX Kidney injury associated molecule; kidney injury related molecule;
 AC KIM; tissue growth promotion; regeneration; renal condition;
 KW acute renal failure; acute nephritis; tumour; ds.
 DT Rattus sp.
 XX Rattus sp.
 OS Rattus sp.
 XX Key Location/Qualifiers
 FH Key complement (42..281)
 FT misc_feature /tag= a
 FT /label= SAC_24409
 FT
 XX W09853071-Al.
 PN W09853071-Al.
 XX 26-NOV-1998.
 PD 26-NOV-1998.
 XX 22-MAY-1998; 98WO-US10547.
 PF 22-MAY-1998; 98WO-US10547.
 XX 23-MAY-1997; 97US-0047491.
 PR 23-MAY-1997; 97US-0047490.
 XX (BIOJ) BIOGEN INC.
 PA Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 PI WPI: 1999-045312/04.
 DR Kidney injury-associated molecule, KIM, polypeptides - upregulated
 XX in injured or regenerating tissues, useful to promote tissue growth
 PT and regeneration, especially to treat renal conditions
 PT
 XX Claim 9; Page 157; 213pp; English.

XX The present sequence represents a kidney injury associated molecule
 CC (KIM) cDNA clone. KIM proteins can be administered therapeutically
 CC by expressing KIM encoding polynucleotides, to promote growth and/or
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
 CC are upregulated in injured or regenerating (especially renal) tissues.
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/disregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically, e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal
 CC tissue growth arising from/affecting renal tissue). The proteins can
 CC also be used to locate KIM-producing cells (especially specific loci,
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
 CC arising from/affecting renal tissue), by contacting cells with an
 CC imageable KIM-binding reagent and imaging reagent accumulation.

XX Sequence 1668 BP; 371 A; 440 C; 422 G; 435 T; 0 other;

Query Match 92.9%; Score 13; DB 20; length 1668;

Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 13
|||||
Db 324 CCTGCTCCCGAGG 312

RESULT 12
AAV72023
ID AAV72023 standard; DNA; 1830 BP.
XX AC
XX AAV72023;
XX 29-MAR-1999 (first entry)
XX ADenovirus Pig4KN DNA.
XX Pig4KN; chimeric protein; adenoviral fibre protein; monomer; infection;
KW trimerisation domain; affinity; substrate; gene therapy vector;
KW attachment; interaction assay; ss.
XX Mastadenovirus.
OS Synthetic.
XX Key Location/Qualifiers
FH 1..1830
FT /*tag= a
FT /product= *Pig4KN"

XX WO9854346-A1.
PN 03-DEC-1998.
XX 28-MAY-1998; 98WO-US11024.
XX 16-JAN-1998; 98US-0071668.
XX 28-MAY-1997; 97US-0047849.
XX (GENV-) GENVEC INC.
XX Brough DE, Einfeld D, Kovesdi I, Lizonova A, Roelvink PW;
PI Wickham TJ, Yonehiro G;
XX WPT: 1999-059848/05.
DR P-PSDB; AAW82727.
XX New adenoviral fibre trimer with reduced binding to native substrate
PT - useful for, e.g. preparing gene therapy vector with minimal
PT ectopic infection for in vitro applications
XX Example 5; Page 52-54; 103pp; English.

XX This sequence encodes a novel adenovirus chimeric protein, Pig4KN. This
CC protein is used in a method for the construction of novel monomers having
CC an N-terminus of an adenoviral fibre protein and a trimerisation domain.
CC Such monomers have lower affinity for native substrate than the native
CC adenoviral fibre trimer. Cell lines containing such monomers are used (i)
CC to propagate adenovirus for use as gene therapy vectors (for in vitro or
CC in vivo applications), (ii) as reagents for studying adenoviral attachment
CC and infection, and (iii) in receptor-ligand interaction assays. The new
CC viruses produce minimal ectopic infection (they can not infect native
CC host cells) so are safer as vectors and can be engineered for selective
CC targeting to other cells.

XX Sequence 1830 BP; 436 A; 586 C; 425 G; 383 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 1830;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 13
|||||
Db 91 cctgctcccgagg 103

RESULT 13
AAF60009
ID AAF60009 standard; DNA; 1954 BP.
XX AC
XX AAF60009;
XX 26-APR-2001 (first entry)
XX Human synthetase #9.
XX Human; synthetase; immune; inflammation; AIDS; infection; cancer;
KW reproduction; ds.
XX Homo sapiens.
OS WO200107628-A2.
PN 01-FEB-2001.
XX 20-JUL-2000; 2000WO-US19980.
XX 22-JUL-1999; 99US-0144992.
PR 02-DEC-1999; 99US-0168858.
XX (INCY-) INCYTE GENOMICS INC.
XX Tang YT, Hillman JL, Bandman O, Yue H, Baughn MR, Lal P, Lu DAM;
PI Shah P, Azimzal Y;
XX WPI; 2001-182799/18.
XX New human synthetases, useful for diagnosing, preventing and treating
PT immune disorders, neuronal disorders, reproductive disorders, and cell
PT proliferative disorders such as cancer -
XX Claim 5; Page 115-116; 120pp; English.

XX The present invention relates to human synthetase proteins. These
CC proteins are useful for treating diseases in which they are involved.
CC Such diseases include immune disorders such as inflammation, acquired
CC immunodeficiency syndrome (AIDS), allergies, autoimmune diseases,
CC infectious cancers, Alzheimer's disease, reproductive disorders and
CC cell proliferative disorders
XX Sequence 1954 BP; 346 A; 702 C; 551 G; 355 T; 0 other;

Query Match 92.9%; Score 13; DB 22; Length 1954;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 13
|||||
Db 1541 cctgctcccgagg 1553

RESULT 14
AAZ36247
ID AAZ36247 standard; cDNA; 1990 BP.
XX AC
XX AAZ36247;
XX 22-FEB-2000 (first entry)

XX cDNA encoding a bone marrow secreted protein designated BMS42.
DE bone marrow secreted protein; bone marrow stromal cell; cytokine;
XX cell proliferation; cell differentiation; hematopoiesis; anaemia;
KW myeloid cell deficiency; lymphoid cell deficiency; myeloid cell;
KW erythroid progenitor cell; colony stimulating factor; granulocyte;
KW monocyte; macrophage; myelo-suppression; megakaryocyte; platelet;
KW platelet disorder; thrombocytopenia; hematopoietic stem cell;

KW stem cell disorder; aplastic anaemia; bone differentiation;
KW paroxysmal nocturnal hemoglobinuria; bone growth; cartilage; tendon;
KW ligament; nerve; wound healing; tissue repair; burn; incision; ulcer;
KW bone fracture; cartilage damage; artificial joint; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 104..1617
FT /*tag= a
FT /product= "bone marrow secreted protein"
FT sig_peptide 104..304
FT /*tag= b
FT polyA_signal 1952..1957
FT /*tag= c
XX
PN WO9933979-A2.
XX
PD 08-JUL-1999.
XX
PF 18-DEC-1998; 98WO-US27008.
XX
PR 30-DEC-1997; 97US-0068958.
PR 24-SEP-1998; 98US-0101603.
PR 30-SEP-1998; 98US-0102540.
XX
PA (CHIR) CHIRON CORP.
XX
PI Lin H, Cao L;
XX
DR WPI; 2000-038344/03.
DR P-PSDB; AAY53641.
XX
PT New isolated human polynucleotide and secreted proteins can induce
PT production of other cytokines in certain cell populations -
XX
PS Claim 11; Page 114-115; 120pp; English.
XX
CC AA236228-49 encode bone marrow secreted proteins of human bone marrow
CC stromal cells. The proteins can exhibit cytokine, cell proliferation, or
CC cell differentiation activity (either inducing or inhibiting). They can
CC be used to support colony forming cells or factor-dependent cell lines,
CC to regulate hematopoiesis, and to treat myeloid or lymphoid cell
CC deficiencies. In addition, they may be used to support the growth and
CC proliferation of erythroid progenitor cells, and to treat various
CC anemias. They can have colony stimulating factor (CSF) activity and can
CC be used to support the growth and proliferation of myeloid cells such as
CC granulocytes, monocytes or macrophages, to prevent or treat
CC myelo-suppression, to support the growth and proliferation of
CC megakaryocytes and platelets, thereby allowing prevention of treatment
CC of platelet disorders such as thrombocytopenia, to support the growth
CC and proliferation of hematopoietic stem cells, either in place of or in
CC conjunction with platelet transfusions, to treat stem cell disorders,
CC such as aplastic anaemia and paroxysmal nocturnal hemoglobinuria, or to
CC repopulate the stem cell compartment after irradiation or chemotherapy.
CC They can be used for growth or differentiation of bone, cartilage,
CC tendon, ligament, or nerve tissue, as well as for wound healing and
CC tissue repair and replacement, and in the treatment of burns, incisions
CC and ulcers, to induce cartilage and/or bone growth in circumstances
CC where bone is not normally formed and thus have an application in healing
CC bone fractures and cartilage damage or defects, prophylactic use in
CC fracture reduction and also in the improved fixation of artificial
CC joints.
XX
SQ Sequence 1990 BP; 344 A; 727 C; 574 G; 345 T; 0 other;

Query Match 92.9%; Score 13; DB 21; Length 1990;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cctgctcccgagg 13
|||||

Db 1574 cctgctcccgagg 1586
RESULT 15
AAV72024
ID AAV72024 standard; DNA; 2253 BP.
XX
AC AAV72024;
XX
DT 29-MAR-1999 (first entry)
XX
DE Adenovirus PNSpig4.SS DNA.
XX
KW PNSpig4.SS; chimeric protein; adenoviral fibre protein; monomer;
KW trimerisation domain; affinity; substrate; gene therapy vector;
KW attachment; interaction assay; infection; ss.
XX
OS Mastadenovirus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..2253
FT /*tag= a
FT /product= "PNSpig4.SS"
FT /note= "Chimeric protein"
XX
PN WO9854346-A1.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US11024.
XX
PR 16-JAN-1998; 98US-0071668.
PR 28-MAY-1997; 97US-0047849.
XX
PA (GENV-) GENVEC INC.
XX
PI Brough DE, Einfeld D, Kovsesdi I, Lizonova A, Roelvink PW;
PI Wickham TJ, Yonehiro G;
XX
DR WPI; 1999-059848/05.
DR P-PSDB; AAV72024.
XX
PT New adenoviral fibre trimer with reduced binding to native substrate
PT - useful for, e.g. preparing gene therapy vector with minimal
PT ectopic infection for in vitro applications
XX
PS Example 6; Page 55-57; 103pp; English.
XX
CC This sequence encodes a novel adenovirus chimeric protein, PNSpig4.SS.
CC This protein is used in a method for the construction of novel monomers
CC having an N-terminus of an adenoviral fibre protein and a trimerisation
CC domain. Such monomers have lower affinity for native substrate than the
CC native adenoviral fibre trimer. Cell lines containing such monomers are
CC used (i) to propagate adenovirus for use as gene therapy vectors (for in
CC vitro or in vivo applications), (ii) as reagents for studying adenoviral
CC attachment and infection, and (iii) in receptor-ligand interaction
CC assays. The new viruses produce minimal ectopic infection (they can not
CC infect native host cells) so are safer as vectors and can be engineered
CC for selective targeting to other cells.
XX
SQ Sequence 2253 BP; 545 A; 735 C; 503 G; 470 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 2253;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cctgctcccgagg 13
|||||
Db 514 cctgctcccgagg 526

us-09-904-420a-5.rng

Thu Feb 21 07:25:58 2002

Search completed: February 20, 2002, 10:31:58
Job time: 10226 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:44:41 : Search time 166.82 Seconds
(without alignments)
19,007 Million cell updates/sec

Title: US-09-904-420A-5
Sequence: 1 cctgctcccgagg 14

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14	100.0	2129	1 US-08-426-819A-34	Sequence 34, Appl
2	14	100.0	2138	3 US-08-776-271-1	Sequence 1, Appl
3	14	100.0	2138	3 US-09-215-035-1	Sequence 1, Appl
4	13	92.9	715	4 US-08-998-416-676	Sequence 676, App
5	13	92.9	152331	3 US-09-128-155-16	Sequence 16, Appl
6	12.4	88.6	27	3 US-09-176-862-22	Sequence 22, Appl
7	12.4	88.6	102	1 US-08-247-475-22	Sequence 22, Appl
8	12.4	88.6	102	1 US-08-479-650-22	Sequence 27, Appl
9	12.4	88.6	102	1 US-08-191-8660-27	Sequence 27, Appl
10	12.4	88.6	102	1 US-08-674-169-22	Sequence 22, Appl
11	12.4	88.6	102	1 US-08-185-9498-27	Sequence 2, Appl
12	12.4	88.6	610	2 US-08-928-926A-2	Sequence 2, Appl
13	12.4	88.6	610	3 US-09-212-149-2	Sequence 2, Appl
14	12.4	88.6	857	5 PCT-US94-05150-19	Sequence 19, Appl
15	12.4	88.6	988	1 US-08-243-545-5	Sequence 5, Appl
16	12.4	88.6	988	2 US-08-993-962-5	Sequence 5, Appl
17	12.4	88.6	988	4 US-09-160-841-5	Sequence 5, Appl
18	12.4	88.6	988	4 US-09-109-100-2	Sequence 2, Appl
19	12.4	88.6	988	5 PCT-US94-05365-5	Sequence 5, Appl
20	12.4	88.6	1057	1 US-08-147-784-1	Sequence 1, Appl
21	12.4	88.6	1057	4 US-08-195-967-1	Sequence 1, Appl
22	12.4	88.6	1057	4 US-08-472-940-1	Sequence 1, Appl
23	12.4	88.6	1176	1 US-08-196-003-1	Sequence 1, Appl
24	12.4	88.6	1176	3 US-08-933-824-1	Sequence 1, Appl
25	12.4	88.6	1176	4 US-09-264-466-1	Sequence 1, Appl
26	12.4	88.6	1230	4 US-09-172-841-52	Sequence 52, Appl
27	12.4	88.6	1756	2 US-08-466-589-3	Sequence 3, Appl

28	12.4	88.6	1756	2 US-08-700-636-3	Sequence 3, Appl
29	12.4	88.6	1756	3 US-08-467-574-3	Sequence 3, Appl
30	12.4	88.6	1756	4 US-09-217-345-3	Sequence 3, Appl
31	12.4	88.6	2126	2 US-08-789-354-1	Sequence 1, Appl
32	12.4	88.6	2126	3 US-09-110-937-1	Sequence 1, Appl
33	12.4	88.6	2126	3 US-09-058-725B-1	Sequence 1, Appl
34	12.4	88.6	2126	3 US-09-232-857-1	Sequence 1, Appl
35	12.4	88.6	2126	1 US-07-736-178C-1	Sequence 1, Appl
36	12.4	88.6	3014	1 US-08-629-939-1	Sequence 1, Appl
37	12.4	88.6	3014	1 US-08-759-873-1	Sequence 1, Appl
38	12.4	88.6	3223	1 US-07-980-528-1	Sequence 36, Appl
39	12.4	88.6	3507	2 US-08-775-009-36	Sequence 1, Appl
40	12.4	88.6	3931	1 US-08-144-121-1	Sequence 1, Appl
41	12.4	88.6	3931	2 US-08-735-893-1	Sequence 18, Appl
42	12.4	88.6	3934	3 US-09-226-568-18	Sequence 1, Appl
43	12.4	88.6	3946	1 US-08-077-848A-1	Sequence 1, Appl
44	12.4	88.6	3946	3 US-09-211-640-1	Sequence 1, Appl
45	12.4	88.6	3946	4 US-09-378-536-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-426-819A-34
; Sequence 34, Application US/08426819A
; Patent No. 5723318
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, No. 5723318omi
; APPLICANT: Kojima, Tetsuo
; APPLICANT: Oh-Eda, Masayoshi
; APPLICANT: Hattori, Kunihiro
; TITLE OF INVENTION: Genes Coding for Megakaryocyte
; TITLE OF INVENTION: Potentiator
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,819A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 230-107P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: pkp027
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..1965
; FEATURE:
; NAME/KEY: misc_feature

;
; LOCATION: 1873
; OTHER INFORMATION: /note= "this residue is A in
; OTHER INFORMATION: PKP021"
US-08-426-819A-34

Query Match 100.0%; Score 14; DB 1; Length 2129;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||
Db 546 CCGTCTCCCGAGG 559

RESULT 2
US-08-776-271-1
; Sequence 1, Application US/08776271
; Patent No. 6083502
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Chang, Kai
; TITLE OF INVENTION: Mesothelin, a Differentiation Antigen
; TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers
; TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,271
; FILING DATE: 01-DEC-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/00224
; FILING DATE: 03-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/010,166
; FILING DATE: 05-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-259100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2138 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 100..1986
US-08-776-271-1

Query Match 100.0%; Score 14; DB 3; Length 2138;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

Db 543 CCGTCTCCCGAGG 556
|||||

RESULT 3
US-09-215-035-1
; Sequence 1, Application US/09215035
; Patent No. 6153430
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Chang, Kai
; TITLE OF INVENTION: Mesothelin, a Differentiation Antigen
; TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers
; TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,035
; FILING DATE: No. 6153430 yet assigned
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/776,271
; FILING DATE: 01-DEC-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/00224
; FILING DATE: 03-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/010,166
; FILING DATE: 05-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-259110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2138 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 100..1986
US-09-215-035-1

Query Match 100.0%; Score 14; DB 3; Length 2138;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||

Db 543 CCGTCTCCCGAGG 556
RESULT 4
US-08-998-416-676
; Sequence 676, Application US/08998416
; Patent No. 6239264

```
;
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgon
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/COC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 676:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1449UP
; US-08-998-416-676

Query Match 92.9%; Score 13; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctgctcccgagg 14
| | | | | | | | | |
Db 535 CTGCTCCGAGG 547

RESULT 5
US-08-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
```

```
;
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 92.9%; Score 13; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctgctcccgagg 14
| | | | | | | | | |
Db 71619 ctgctcccgagg 71631

RESULT 6
US-09-176-862-22
; Sequence 22, Application US/09176862B
; Patent No. 6046319
; GENERAL INFORMATION:
; APPLICANT: Power, Christopher
; APPLICANT: Mayne, Michael B.
; TITLE OF INVENTION: ANTISENSE OLIGODEOXYNUCLEOTIDES REGULATING EXPRESSION
; FILE REFERENCE: 3045.00002
; CURRENT APPLICATION NUMBER: US/09/176,862B
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: 60/062,718
; EARLIER FILING DATE: 1997-10-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
US-09-176-862-22

Query Match 88.6%; Score 12.4; DB 3; Length 27;
Best Local Similarity 92.9%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
| | | | | | | | | |
Db 8 cctgctcccgagg 21

RESULT 7
US-08-247-475-22
; Sequence 22, Application US/08247475
; Patent No. 5593873
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Macdonald, Richard D.
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
```


; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,475
; FILING DATE: May 23, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-247-475-22

Query Match 88.6%; Score 12.4; DB 1; Length 102;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 cctgctcccgagg 14
|||||
Db 66 CCTGCTCCGACGG 79

RESULT 8
US-08-479-650-22
; Sequence 22, Application US/08479650
; Patent No. 5599544
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Macdonald, Richard D.
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,650
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-479-650-22

Query Match 88.6%; Score 12.4; DB 1; Length 102;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 cctgctcccgagg 14
|||||
Db 66 CCTGCTCCGACGG 79

RESULT 9
US-08-191-866D-27
; Sequence 27, Application US/08191866D
; Patent No. 5783195
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Macdonald, Richard D.
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/191,866D
; FILING DATE: 4 February 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-191-866D-27

Query Match 88.6%; Score 12.4; DB 1; Length 102;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 cctgctcccgagg 14
|||||
Db 66 CCTGCTCCGACGG 79

RESULT 10

```
US-08-674-169-22
; Sequence 22, Application US/08674169
; Patent No. 5804372
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Macdonald, Richard D.
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,169
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-674-169-22

Query Match 88.6%; Score 12.4; DB 1; Length 102;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
   |||||
Db 66 CCGTCTCCCGACGG 79

RESULT 11
US-08-185-949B-27
; Sequence 27, Application US/08185949B
; Patent No. 5874279
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; APPLICANT: Richard D. Macdonald
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2

US-08-674-169-22
; Sequence 22, Application US/08674169
; Patent No. 5804372
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Macdonald, Richard D.
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,169
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-674-169-22

Query Match 88.6%; Score 12.4; DB 1; Length 102;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
   |||||
Db 66 CCGTCTCCCGACGG 79

RESULT 12
US-08-928-926A-2/c
; Sequence 2, Application US/08928926A
; Patent No. 5925543
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NADH DEHYDROGENASE B17 SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,926A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0384 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 base pairs
```

us-09-904-420a-5.rni

Thu Feb 21 07:25:59 2002

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TBLYN0T01
; CLONE: 44898
; US-08-928-926A-2

Query Match      88.6%; Score 12.4; DB 2; Length 610;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 cctgctcccgagg 14
      || |||||
Db      210 CCGGCTCCGAGGG 197

RESULT 13
US-09-212-149-2/c
; Sequence 2, Application US/09212149
; Patent No. 6100036
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NADH DEHYDROGENASE B17 SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/212,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,926
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0384 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TBLYN0T01
; CLONE: 44898
; US-09-212-149-2

Query Match      88.6%; Score 12.4; DB 3; Length 610;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 cctgctcccgagg 14
      || |||||
Db      210 CCGGCTCCGAGGG 197

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TBLYN0T01
; CLONE: 44898
; US-08-928-926A-2

Query Match      88.6%; Score 12.4; DB 2; Length 610;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 cctgctcccgagg 14
      || |||||
Db      210 CCGGCTCCGAGGG 197

RESULT 14
PCT-US94-05150-19/c
; Sequence 19, Application PC/TUS9405150
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Purified Mammalian Flt3 Ligands and Agonists and Antagonist
; NUMBER OF SEQUENCES: 37
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05150
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/162,413
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,111
; FILING DATE: 19-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,391
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,340
; FILING DATE: 13-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/092,549
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,263
; FILING DATE: 07-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,231
; FILING DATE: 19-MAY-1993
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US94-05150-19

Query Match      88.6%; Score 12.4; DB 5; Length 857;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 cctgctcccgagg 14
      || |||||
Db      751 CCTGCTCCCGAGGG 738

RESULT 15
US-08-243-545-5/c
; Sequence 5, Application US/08243545
; Patent No. 5554512
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; TITLE OF INVENTION: Ligands for flt3/flk-2 Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Stephen I. Malaska, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
```

Thu Feb 21 07:25:59 2002

COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0.1
SOFTWARE: Microsoft Word, Version #5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/243,545
FILING DATE: 11-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/162,407
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: 08/111,758
FILING DATE: August 25, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/106,463
FILING DATE: August 12, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,394
FILING DATE: May 24, 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Malaska, Stephen L.
REGISTRATION NUMBER: 32,655
REFERENCE/DOCKET NUMBER: 2813-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 30..734
US-08-243-545-5

Query Match 88.6%; Score 12.4; DB 1; Length 988;
Best Local Similarity 92.9%; Pred. No. 2.6e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 cctgctcccgagg 14
| | | | | | | | | | | | | | | |
Db 690 CTTGCTCCCGAGG 677

Search completed: February 20, 2002, 07:44:49
Job time: 197 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:25:41 ; Search time 3814.84 Seconds
(without alignments)
39.436 Million cell updates/sec

Title: US-09-904-420a-5
Sequence: 1 cctgctccgaggg 14

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estom:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estov:*
8:	em_estov:*
9:	em_htc:*
10:	gb_estl:*
11:	gb_estl:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_pro:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	14	100.0	121	AA275661	AA275661 vc33a09.1
C 2	14	100.0	176	AA765168	AA765168 n279c08.s
C 3	14	100.0	184	AA731084	AA731084 n268a03.s
C 4	14	100.0	234	AA293106	AA293106 AV293106
C 5	14	100.0	273	AA311787	AA311787 EST182503
C 6	14	100.0	281	AA325590	AA325590 17905 MAR
C 7	14	100.0	283	BB607667	BB607667 BB607667
C 8	14	100.0	288	BB367350	BB367350 BB367350
C 9	14	100.0	337	B1403966	B1403966 MI-P-CP1-
C 10	14	100.0	347	BE227885	BE227885 894035H09
C 11	14	100.0	353	BC386035	BC386035 602455204
C 12	14	100.0	354	BE092997	BE092997 CW1-BT074

13	14	100.0	355	11	F14893
14	14	100.0	388	11	BF940593
15	14	100.0	405	10	AM463205
16	14	100.0	418	11	BF940615
17	14	100.0	442	10	AM975810
18	14	100.0	459	11	BF194763
19	14	100.0	498	10	A1951520
20	14	100.0	500	10	AM659550
21	14	100.0	503	13	AG937704
22	14	100.0	504	13	A2617446
23	14	100.0	531	11	BF434130
24	14	100.0	539	10	A1673528
25	14	100.0	542	10	AL556629
26	14	100.0	565	10	BE732462
27	14	100.0	566	11	BF127436
C 27	14	100.0	598	13	A2999674
C 28	14	100.0	598	13	A2999674
29	14	100.0	610	11	B1333483
30	14	100.0	619	13	FR0005293
31	14	100.0	650	10	BE376302
32	14	100.0	678	11	B1185132
C 33	14	100.0	684	10	BE278765
C 34	14	100.0	699	10	AL556126
C 35	14	100.0	724	10	AU133492
36	14	100.0	724	11	BE900825
C 37	14	100.0	748	11	BE917972
38	14	100.0	778	10	BE286940
39	14	100.0	780	11	BE903130
40	14	100.0	788	11	BE251659
41	14	100.0	817	10	AL555912
42	14	100.0	825	10	AL554741
43	14	100.0	847	13	A2704040
44	14	100.0	912	10	AL514070
C 45	14	100.0	913	13	CNS03100

ALIGNMENTS

RESULT 1
AA275661/c

LOCUS: AA275661 121 bp mRNA EST 01-APR-1997
DEFINITION: VC33a09.r1 Barstead MPLRB1 Mus musculus CDNA clone IMAGE:776344 5' similar to SW:COX1_DIDMA P41310 CYTOCHROME C OXIDASE POLYPEPTIDE 1 ;, mRNA sequence.

ACCESSION: AA275661 GI:1918165
VERSION: AA275661
KEYWORDS: EST.
SOURCE: house mouse.

ORGANISM: Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE: 1 (bases 1 to 121)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE: The WashU-HMMI Mouse EST Project
JOURNAL: Unpublished (1996)
COMMENT: Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INLNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:469200

Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 1.

F14893 SSC1D08 Por
BF940593 nae22f10
AM463205 BF230012A
BF940615 nae23a04
AM975810 EST387919
BF194763 7093c04.x
A1951520 wv36a10.x
AM659550 96847 MAR
AG937704 NB6-335R
A2617446 1M0448B15
BF434130 7099005.x
A1673528 wv5004.x
AL556629 AL556629
BE732462 601567623
BF127436 601810043
A2999674 2M0287C11
B1333483 602996934
Z89103 F.rubripes
BE376302 60128671
B1185132 UNI-P-FN-
BE278765 601158985
AL556126 AL556126
AU133492 AU133492
BE900825 601674556
BE917972 602818907
BE286940 601092425
BE903130 601673130
BE251659 602363552
AL555912 AL555912
AL554741 AL554741
A2704040 RPT-23-2
AL514070 AL514070
AL245817 Tetraodon

FEATURES
source

Location/Qualifiers
1. .121
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="IMAGE:776344"
/clone_lib="Barstead MPLRBI"
/sex="mixed"
/tissue_type="Kidney"
/dev_stage="6 weeks"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'-
TGTTACGAATCTGAAGTGGAGCGCGCCCTCTTTTTTTTTTTT-3'];
double-stranded cDNA was ligated to Eco RI adaptors
[CATGATTCGGTACCT], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."

BASE COUNT
ORIGIN

37 a 35 c 22 g 27 t

Query Match 100.0%; Score 14; DB 10; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

|||||

Db 96 cctgctcccgagg 83

RESULT 2

AA765168 176 bp mRNA EST 07-FEB-1998
DEFINITION n279c08.sl NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1301678 3',
mRNA sequence.
AA765168
ACCESSION AA765168
VERSION AA765168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1282 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 173.

FEATURES
source

Location/Qualifiers
1. .176
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1301678"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, Igp-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACGAATCTGAAGTGGAGCGCGCCCTCTTTTTTTTTTTT-3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 34 a 47 c 59 g 36 t

ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14

|||||

Db 119 cctgctcccgagg 132

RESULT 3

AA731084 184 bp mRNA EST 07-FEB-1998
DEFINITION nw68a03.sl NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251724 3',
mRNA sequence.
AA731084
ACCESSION AA731084
VERSION AA731084
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1310 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 183.

FEATURES
source

Location/Qualifiers
1. .184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1251724"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, Igp-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACGAATCTGAAGTGGAGCGCGCCCTCTTTTTTTTTTTT-3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```

and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
34 a 49 c 63 g 37 t 1 others
BASE COUNT
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 184;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
Db 118 CCTGCTCCCGAGG 131

RESULT 4
AV293106/c
LOCUS
DEFINITION AV293106 RIKEN full-length enriched, 6 days neonate head Mus
musculus cDNA clone 5430433E23 3', mRNA sequence.
ACCESSION AV293106
VERSION AV293106.1 GI:6307137
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 234)
Konno,H., Aizawa,K., Akahira,S., Akiyama,S., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
Y., Shigenoto,Y., Shiaki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome.res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source Location/Qualifiers
1..234
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="5430433E23"
/head"

/sex="mixed"
/tissue_type="head"
/dev_stage="6 days neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCCAAGACTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATTCCTCCCTCCCTCC
3'], cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC 1. Cloning sites, 5' end: Sali; 3' end:
BamHI."
BASE COUNT 44 a 63 c 62 g 65 t
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
Db 24 CCTGCTCCCGAGG 11

RESULT 5
AA311787/c
LOCUS
DEFINITION EST182503 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
similar to T-cell factor 1, A/B/C, mRNA sequence.
ACCESSION AA311787
VERSION AA311787.1 GI:1964114
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 273)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,
Ghosh,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,A.,
Kelley,J.C., Liu,L.-I., Marnaros,S.M., Merrick,J.M., Pelligrino,S.M.,
Moreno-Palaoques,R.F., McDonald,L.A., Nguyen,D.T., Shirley,R.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.P., Fertie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olson,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Pannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA

```

Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tldb/hgi/hgi.html>)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1. .273

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="ATCC (inhost):158998"
 /db_xref="taxon:9606"
 /clone_lib="Jurkat T-cells VI"
 /cell_type="T-lymphocyte"
 /note="Vector: pBluescript SK-; Site_1: ECORI; Site_2:
 XhoI"

BASE COUNT 60 a 111 c 59 g 39 t 4 others
 ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 273;
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14
 |||||
 Db 153 CCGTCTCCGAGG 140

RESULT 6
 AW325590 281 bp mRNA EST 09-JUL-2000
 LOCUS 17905 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION AW325590
 ACCESSION AW325590.1 GI:6761511
 VERSION
 KEYWORDS EST.
 SOURCE pig.

ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 281)
 Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
 and Keele,J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine

Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 20
 and -minmatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGCAGG
 Plate: 8 row: 8 column: 8
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1. .281

FEATURES
 source

/organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC lPIG"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 73 a 85 c 82 g 41 t
 ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 281;
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14
 |||||
 Db 200 CCGTCTCCGAGG 213

FEATURES
 source

RESULT 7
 BB607667/c 283 bp mRNA EST 06-DEC-2000
 LOCUS BB607667 RIKEN full-length enriched, 2 days pregnant adult female
 DEFINITION oviduct Mus musculus cDNA clone E230006M17 5', mRNA sequence.

ACCESSION BB607667
 VERSION BB607667.1 GI:11561533
 KEYWORDS EST.
 SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 283)
 Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.,
 Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodozawa,Y.,
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
 H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
 Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
 Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
 Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
 T., Toya,T., Wataniki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
 Unpublished (2000)

TITLE Contact: Yoshihide Hayashizaki
 JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic
 COMMENT Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

URL: <http://genome.res.gsc.riken.go.jp/>
 Email: genome-res@gsc.riken.go.jp/
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
 Thermostabilization and thermoactivation of full length
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Kitzunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.
 Location/Qualifiers
 1. .283

FEATURES
 source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="E230006M17"
 /clone_lib="RIKEN full-length enriched, 2 days pregnant
 adult female oviduct"
 /sex="female"
 /tissue_type="oviduct"
 /dev_stage="2 days pregnant adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCAACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCTCGAGTTAAATTAATCCCGCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

75 a 64 c 80 g 64 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||

Db 163 cctgctcccgagg 150

RESULT 8
BB367350/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BB367350 288 bp mRNA EST 12-JUL-2000
musculus cDNA clone Cl30037J03 3', mRNA sequence.
BB367350
EST.
house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 288)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

TITLE
JOURNAL
COMMENT

FEATURES
source

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

1..288
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="Cl30037J03"
head
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCAACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCTCGAGTTAAATTAATCCCGCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 37 a 100 c 96 g 55 t

ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctgctcccgagg 14
|||||

Db 157 cctgctcccgagg 144

RESULT 9

BI403966/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Tugle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktugle@iastate.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the Nott site

and the oligo-dT track served to verify it as a clone from the normalized uterus cDNA Library preparation: M.B. Soares Lab , University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers

1. .337
/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-CPI-nwr-g-11-0-UI"
/clone_lib="MI-P-CPI"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT730-Pac (Pharmacia) with a modified polylinker: Site1: Not I; Site2: EcoRI; The MI-P-CPI library is normalized library derived from the MI-P-CP0 library, ultimately derived from uterus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iasatate.edu/>. The procedure used to create this library has been previously described (Bonaldo , Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_LIB=MI-P-CPI
TAG_TISSUE=uterus
TAG_SEQ=AGTCCAATCG"

BASE COUNT 56 a 86 c 91 g 103 t 1 others

ORIGIN

Query Match 100.0%; Score 14; DB 11; Length 337;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14

Db 242 CCTGCTCCCGAGG 229

RESULT 10

LOCUS BE227885 347 bp mRNA EST 06-JUL-2000
DEFINITION 894035H09.xl C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION

BE227885

VERSION

BE227885.1 GI:8933124

KEYWORDS

EST.

SOURCE

Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
Unpublished (2000)

COMMENT

Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.

FEATURES

Location/Qualifiers

1. .347
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site1: EcoRI; Site2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 65 a 121 c 92 g 69 t

ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 347;

Best Local Similarity 100.0%; Pred. No. 2.5e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgctcccgagg 14

Db 314 CCTGCTCCCGAGG 327

RESULT 11

EG386035/c

LOCUS

602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5', mRNA EST 12-MAR-2001

DEFINITION

EG386035

ACCESSION

EG386035

VERSION

EG386035.1 GI:13279481

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 353)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-f@mail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCMI307 row: h column: 02

High quality sequence stop: 276.

Location/Qualifiers

FEATURES

source

1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site1: XhoI; Site2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(C). Size selected >500bp for average insert size 1.8Kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)";

BASE COUNT

80 a 111 c 122 g 40 t

ORIGIN

Query Match	100.0%;	Score 14;	DB 11;	Length 353;	
Best Local Similarity	100.0%;	Pred. No. 2.5e+03;			
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	cctgctcccgagg 14			
Db	269	CCTGCTCCCGAGGG 256			
RESULT 12					
BE092997/c					
LOCUS		354 bp	mRNA	EST	12-JUN-2000
DEFINITION		CM1-BT0742-020500-211-e08	BT0742	Homo sapiens	cdNA, mRNA sequence.
ACCESSION		BE092997			
VERSION		BE092997.1	GI:8483449		
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE					
AUTHORS		1 (bases 1 to 354)			
		Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.			
TITLE		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)			
MEDLINE		20202663			
COMMENT		Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=st2-CM1-BT0742-020500-211-e08&t3=2000-05-02&t4=1) Seq primer: puc 18 forward High quality sequence start: 16 High quality sequence stop: 354.			
FEATURES					
source		1..354			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone_lib="BT0742"			
		/dev_stage="Adult"			
		/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT		98 a	89 c	79 g	88 t
ORIGIN					
Query Match	100.0%;	Score 14;	DB 10;	Length 354;	
Best Local Similarity	100.0%;	Pred. No. 2.5e+03;			
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	cctgctcccgagg 14			
Db	56	CCTGCTCCCGAGGG 43			
RESULT 13					

F14893					
LOCUS		355 bp	mRNA	EST	09-SEP-1996
DEFINITION		SSC1D08 Porcine small intestine	cdNA library	Sus scrofa	cdNA clone
ACCESSION		F14893			
VERSION		F14893.1	GI:971724		
KEYWORDS		EST.			
SOURCE		pig.			
ORGANISM		Sus scrofa			
REFERENCE					
AUTHORS		1 (bases 1 to 355)			
TITLE		Wintero, A.K., Fredholm, M. and Davies, W.			
		Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones			
JOURNAL		Mamm. Genome 7 (7), 509-517 (1996)			
MEDLINE		96327607			
COMMENT		Contact: A.K. Wintero Department of Animal Science and Animal Health, Division of Animal Genetics, The Royal Veterinary and Agricultural University Bulowsvej 13, 1870 Frederiksberg C, Denmark.			
FEATURES					
source		1..355			
		/organism="Sus scrofa"			
		/db_xref="taxon:9823"			
		/clone_lib="cd08"			
		/note="Porcine small intestine cDNA library"			
		/note="directionally cloned cDNA in X11-blue MRF"			
BASE COUNT		64 a	102 c	102 g	81 t
ORIGIN					
Query Match	100.0%;	Score 14;	DB 11;	Length 355;	
Best Local Similarity	100.0%;	Pred. No. 2.5e+03;			
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	cctgctcccgagg 14			
Db	232	CCTGCTCCCGAGGG 245			
RESULT 14					
BE940593					
LOCUS		388 bp	mRNA	EST	30-MAR-2001
DEFINITION		nae22110.x1 NCI-CCAP_Ov18	Homo sapiens	cdNA clone	IMAGE:3436171 3'
		similar to contains TAR1.t3	TAR1	repetitive element	;; mRNA
ACCESSION		BF940593			
VERSION		BF940593.1	GI:12357913		
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE					
AUTHORS		1 (bases 1 to 388)			
TITLE		NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.			
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
JOURNAL		Unpublished (1997)			
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov Seq primer: -400P from Gibco High quality sequence stop: 321. Location/Qualifiers			
FEATURES					
source		1..388			

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3436171"
/clone_lib="NCL.CGAP.Ovl8"
/tissue_type="fibroblasts"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGCGGCGGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 50 a 131 c 165 g 296 t
ORIGIN

Query Match 100.0%; Score 14; DB 11; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctcccgagg 14
 | | | | | | | | | |
DB 227 CCTGCTCCGAGG 240

RESULT 15
AW463205
LOCUS
DEFINITION
clone BP230012A10B5 5', mRNA sequence.
ACCESSION AW463205
VERSION AW463205.1 GI:7033373
KEYWORDS
SOURCE EST.
ORGANISM
Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 405)
Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
J.H.
Bovine ESTs
Unpublished (2000)
Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimming g:
Cross-match from Washington University Genome Center PHRAP suite.
Sequences submitted are vector free and at least 200 bp in length.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCCCTCACTAAAG
Insert Length: 405 Std Error: 0.00
Plate: BP230012A10 row: B column: 5
Seq primer: AGCGATAACAATTCACAGGA
High quality sequence stop: 405.
Location/Qualifiers
1. .405
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP230012A10B5"

FEATURES
source
1. .405
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP230012A10B5"
/clone_lib="Soares normalized bovine placenta"

/sex="female"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI; The cDNA library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806."
BASE COUNT 86 a 112 c 108 g 96 t
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 405;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtctcccgagg 14
 | | | | | | | | | |
DB 289 CCTGCTCCGAGG 302

Search completed: February 20, 2002, 10:25:45
Job time: 9853 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 20, 2002, 07:41:32 ; Search time 1729.17 Seconds
(without alignments)
162.189 Million cell updates/sec

Title: US-09-904-420A-4
Perfect score: 17
Sequence: 1 catcattaccaccacca 17
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vi.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	100.0	179641	2	AC073694	AC073694 Mus muscu
2	17	100.0	186243	2	AC073740	AC073740 Mus muscu
3	16	94.1	1115	9	HSTCF1D	247361 H.sapiens T
4	16	94.1	1165	9	HSTCF1E	247362 H.sapiens T
5	16	94.1	1254	9	HSTCF1A	X59869 Human TCF-1
6	16	94.1	2814	9	HSTCF1B	X59870 Human TCF-1
7	16	94.1	2855	9	HSTCF1G	X63901 Homo sapien
8	16	94.1	2910	9	HSTCF1C	X59871 Human TCF-1
9	16	94.1	84544	2	AC009012	AC009012 Homo sapi
10	16	94.1	86914	2	AC011336	AC011336 Homo sapi
11	16	94.1	159420	2	AC009017	AC009017 Homo sapi
12	16	94.1	177562	9	HS269M15	AL021395 Human DNA
13	16	94.1	200831	2	AC008608	AC008608 Homo sapi
14	15.4	90.6	216	11	HSPF15G5	294649 H.sapiens f
15	15.4	90.6	2046	9	AF097994	AF097994 Homo sapi
16	15.4	90.6	33478	3	CEP53B2	Z73908 Caenorhabdi
17	15.4	90.6	36759	2	AC019705	AC019705 Drosophil
18	15.4	90.6	93626	2	AC010042	AC010042 Drosophil
19	15.4	90.6	95091	9	AL353594	AL353594 Human DNA
20	15.4	90.6	99509	3	LMFP1295	AL359773 Leishmani
21	15.4	90.6	102552	2	AC019368	AC019368 Drosophil
22	15.4	90.6	117609	9	AL354773	AL354773 Human DNA
23	15.4	90.6	120773	9	AL139322	AL139322 Human DNA
24	15.4	90.6	125013	2	DMBR11110	AL122028 Drosophil
25	15.4	90.6	134594	2	AC083822	AC083822 Homo sapi
26	15.4	90.6	138230	2	AL391063	AL391063 Homo sapi
27	15.4	90.6	138633	2	AL356360	AL356360 Homo sapi
28	15.4	90.6	149896	2	AC026467	AC026467 Homo sapi
29	15.4	90.6	153108	9	AL390195	AL390195 Human DNA
30	15.4	90.6	154011	2	AC018571	AC018571 Homo sapi
31	15.4	90.6	157709	2	AC090262	AC090262 Homo sapi
32	15.4	90.6	161271	9	AC011200	AC011200 Homo sapi
33	15.4	90.6	162099	2	AC015730	AC015730 Homo sapi
34	15.4	90.6	166995	2	AC022631	AC022631 Homo sapi
35	15.4	90.6	169832	2	AL590669	AL590669 Homo sapi
36	15.4	90.6	171905	2	AL591609	AL591609 Homo sapi
37	15.4	90.6	172945	9	AC007220	AC007220 Homo sapi
38	15.4	90.6	176226	2	AC026583	AC026583 Homo sapi
39	15.4	90.6	177989	2	AL591430	AL591430 Mus muscu
40	15.4	90.6	180298	2	AL357633	AL357633 Homo sapi
41	15.4	90.6	184562	9	AC016575	AC016575 Homo sapi
42	15.4	90.6	185243	2	AC021958	AC021958 Homo sapi
43	15.4	90.6	190982	2	AC025285	AC025285 Homo sapi
44	15.4	90.6	191583	9	AC015651	AC015651 Homo sapi
45	15.4	90.6	196915	2	AC046185	AC046185 Homo sapi

ALIGNMENTS

RESULT 1
AC073694
LOCUS AC073694 179641 bp DNA 18-JUL-2000
DEFINITION Mus musculus clone RP23-159B10, WORKING DRAFT SEQUENCE, 21 ordered pieces.
ACCESSION AC073694
VERSION AC073694.2 GI:9256760
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 179641)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179641)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint

COMMENT

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810311.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1797962
Center clone name: RPCI-23_159B10

Summary Statistics
Consensus quality: 169003 bases at least Q40
Consensus quality: 175721 bases at least Q30
Consensus quality: 177040 bases at least Q20
Estimated insert size: 182000; agarose-fp estimation
Estimated insert size: 178691; sum-of-contigs estimation
Quality coverage: 8.85 in Q20 bases; agarose-fp estimation
Quality coverage: 9.01 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 9983: contig of 9983 bp in length
* 9984 10083: gap of unknown length
* 10084 24408: contig of 14325 bp in length
* 24409 24508: gap of unknown length
* 24509 36878: contig of 12370 bp in length
* 36879 36978: gap of unknown length
* 36979 44353: contig of 7375 bp in length
* 44354 44454: gap of unknown length
* 44454 47523: contig of 3070 bp in length
* 47524 47624: gap of unknown length
* 47624 53213: contig of 5590 bp in length
* 53214 53313: gap of unknown length
* 53314 58877: contig of 5564 bp in length
* 58878 58977: gap of unknown length
* 58978 60867: contig of 1890 bp in length
* 60868 60967: gap of unknown length
* 60968 70337: contig of 9370 bp in length
* 70338 70437: gap of unknown length
* 70438 72640: contig of 2203 bp in length
* 72641 72740: gap of unknown length
* 72741 75634: contig of 2894 bp in length
* 75635 75734: gap of unknown length
* 75735 78720: contig of 2986 bp in length
* 78721 78820: gap of unknown length
* 78821 84547: contig of 5727 bp in length
* 84548 84647: gap of unknown length
* 84648 115904: contig of 31257 bp in length
* 115905 116004: gap of unknown length
* 116005 141315: contig of 25311 bp in length
* 141316 141415: gap of unknown length
* 141416 151262: contig of 9847 bp in length
* 151263 151362: gap of unknown length
* 151363 166484: contig of 15122 bp in length
* 166485 166584: gap of unknown length
* 166585 169891: contig of 3307 bp in length
* 169892 169991: gap of unknown length
* 169992 172410: contig of 2419 bp in length
* 172411 172510: gap of unknown length
* 172511 178861: contig of 6351 bp in length
* 178862 178961: gap of unknown length
* 178962 179641: contig of 680 bp in length.

FEATURES
source

1. .179641
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT
ORIGIN

/clone="RP23-159B10"
/clone_lib="RPCI mouse BAC library 23"
44967 a 46477 c 44480 g 41715 t 2002 others

Query Match 100.0%; Score 17; DB 2; Length 179641;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgattaccaccacca 17
|||||

Db 156342 CATGCATTACCACCCA 156358

RESULT 2

AC073740/c

LOCUS

DEFINITION

AC073740 186243 bp DNA HTG 18-JUL-2000

Mus musculus clone RP23-274122, WORKING DRAFT SEQUENCE, 11 ordered

pieces

AC073740

AC073740.2 GI:9256778

HTG; HTGS_PHASE2; HTGS_DRAFT.

house mouse.

Mus musculus

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 186243)
Sequencing of Mouse
Unpublished
2 (bases 1 to 186243)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810357.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 1842302

Center clone name: RPCI-23_274122

Summary Statistics

Consensus quality: 175948 bases at least Q40
Consensus quality: 182693 bases at least Q30
Consensus quality: 184188 bases at least Q20
Estimated insert size: 202290; agarose-fp estimation
Estimated insert size: 185793; sum-of-contigs estimation
Quality coverage: 6.58 in Q20 bases; agarose-fp estimation
Quality coverage: 7.17 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 438: contig of 438 bp in length
* 439 538: gap of unknown length
* 539 12616: contig of 12078 bp in length
* 12617 12716: gap of unknown length
* 12717 24949: contig of 12233 bp in length
* 24950 25049: gap of unknown length
* 25050 42617: contig of 17568 bp in length
* 42618 92550: contig of 49833 bp in length
* 92551 92650: gap of unknown length

* 92651 94922: contig of 2272 bp in length
 * 94923 95022: gap of unknown length
 * 95023 97020: contig of 1998 bp in length
 * 97021 97120: gap of unknown length
 * 97121 157072: contig of 5952 bp in length
 * 157073 157172: gap of unknown length
 * 157173 174862: contig of 17690 bp in length
 * 174863 174962: gap of unknown length
 * 174963 178440: contig of 3478 bp in length
 * 178441 178540: gap of unknown length
 * 178541 186243: contig of 7703 bp in length.

FEATURES

source

Location/Qualifiers
 1. .186243
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-274122"
 /clone_lib="RPC1 mouse BAC library 23"
 BASE COUNT 43755 a 47751 c 48436 g 45300 t 1001 others
 ORIGIN

Query Match 100.0%; Score 17; DB 2; Length 186243;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgattaccacca 17
 |||||
 Db 171731 CATGCATTACCCACCA 171715

RESULT 3

HSTCF1D HSTCF1D 1115 bp mRNA PRI 09-JAN-1995
 LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.
 DEFINITION
 ACCESSION Z47361
 VERSION Z47361.1 GI:619881
 KEYWORDS splice form D; T cell factor 1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1115)
 REFERENCE
 AUTHORS Mayer,K., Wolff,E., Clevers,H. and Ballhausen,W.G.
 TITLE The human high mobility group (HMG)-box transcription factor TCF-1:
 novel isoforms due to alternative splicing and usage of a new exon
 IXA

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1115)
 AUTHORS van de Wetering,M., Oosterwegel,M., Holstege,F., Dooyes,D.,
 Suijkerbuijk,R., Geurts van Kessel,A. and Clevers,H.
 TITLE The human T cell transcription factor-1 gene. Structure,
 localization, and promoter characterization
 JOURNAL J. Biol. Chem. 267 (12), 8530-8536 (1992)
 MEDLINE 92235082
 REFERENCE 3 (bases 1 to 1115)
 AUTHORS Ballhausen,W.G.
 TITLE Direct Submission
 JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
 Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
 Schwabachanlage 10, Erlangen, Germany, D-91054
 Location/Qualifiers
 1. .1115

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="lymphoma"
 /cell_type="T-lymphocyte"
 /cell_line="Jurkat"
 1. .1115
 /gene="TCF-1"
 /standard_name="T cell factor 1 splice form D"
 /citation=[1]
 /function="high mobility group box transcription factor"
 /evidence=experimental

mRNA

gene 1. .1115
 /gene="TCF-1"
 CDS 2. .754
 /gene="TCF-1"
 /standard_name="T cell factor 1 splice form D"
 /citation=[1]
 /codon_start=1
 /product="T cell factor 1 splice form D"
 /protein_id="CAA87439.1"
 /db_xref="GI:619882"
 /translation="MYKETVYSAFNLLMYPPPSGAGHPQPPPLHKANOPPHGVPO
 LSLYHFNSPHPTAPADISQKVHRPLOTDLGSFYSLTSGMQLPHTVSWFTHPS
 LMLGSGVPGHPAAIPHPAIVPPSGKQELQDFDRNLKTAESKAEREAKPTIKKPLNA
 FMLYKEMRAKVIAECTLKESAAINQILGRWHALSKSEQAKYVELARKERQLHMQLY
 PGWSARDNYGKKRRSRSEKHQESTTDSNLHYS"
 731. .751
 mat_peptide
 /gene="TCF-1"
 /product="alternative ORF specific for TCF-1b"
 BASE COUNT 272 a 407 c 265 g 171 t
 ORIGIN

Query Match 94.1%; Score 16; DB 9; Length 1115;
 Best Local Similarity 100.0%; Pred. No. 7e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgattaccaccc 16
 |||||
 Db 40 CATGCATTACCCACCC 55

RESULT 4

HSTCF1E HSTCF1E 1165 bp mRNA PRI 09-JAN-1995
 LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.
 DEFINITION
 ACCESSION Z47362
 VERSION Z47362.1 GI:619883
 KEYWORDS splice form E; T cell factor 1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1165)
 REFERENCE
 AUTHORS Mayer,K., Wolff,E., Clevers,H. and Ballhausen,W.G.
 TITLE The human high mobility group (HMG)-box transcription factor TCF-1:
 novel isoforms due to alternative splicing and usage of a new exon
 IXA

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1165)
 AUTHORS van de Wetering,M., Oosterwegel,M., Holstege,F., Dooyes,D.,
 Suijkerbuijk,R., Geurts van Kessel,A. and Clevers,H.
 TITLE The human T cell transcription factor-1 gene. Structure,
 localization, and promoter characterization
 JOURNAL J. Biol. Chem. 267 (12), 8530-8536 (1992)
 MEDLINE 92235082
 REFERENCE 3 (bases 1 to 1165)
 AUTHORS Ballhausen,W.G.
 TITLE Direct Submission
 JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer
 Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg,
 Schwabachanlage 10, Erlangen, Germany, D-91054
 Location/Qualifiers
 1. .1165

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="lymphoma"
 /cell_type="T-lymphocyte"
 /cell_line="Jurkat"
 1. .1165
 /gene="TCF-1"
 /standard_name="T cell factor 1 splice form E"
 /citation=[1]
 /function="high mobility group box transcription factor"

```

/evidence=experimental
1. .1165
/gene="TCF-1"
CDS
2. .1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKTVYSAFNLLMHYPSPGAGOHPOPLHKANOPPHGVPO
LSLYEHFNSPHPTAPADISQKVHRPLOTDLSGFISLTSGSMGQLPHTVSWFTHPS
LMLGSGVPGHPAIPAIPAIYPPSGKOELOPDRNLKTOAESKAEEKKPTIKKPLNA
FMLYMKEAKRAVIAECTLKESAAINQILGRWHALSREFOAKYELARKERQLHMOLY
PGMSARDNYGKKRRSRKREHQSSTDPGSKRCARFGLNQOTDMCGPCRRKKKCIKY
LPGEGPCPSVSDSALGPGSPAQDQSPSYHLIPREPTLLTSPAERHLHPQVSP
LSASQPGQHPRPPAAPCRAHRYSNRNLDRWPSRHTPGRLQEP"
803. .1141
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN

Query Match 94.1%; Score 16; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 6.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccaccacc 16
|||||
Db 40 CATGCATTACCCACCC 55

RESULT 5
HSTCF1A HSTCF1A 1254 bp mRNA PRI 14-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form A).
DEFINITION X59869 X55327
ACCESSION X59869.1 GI:36785
VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES van de Wetering, M.
JOURNAL Direct Submission.
Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
REFERENCE 2 (bases 1 to 1254)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H.
TITLES Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box
JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 91114695
COMMENT See also X59869-X59871.
FEATURES
source
Location/Qualifiers
1. .1254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat"
/clone_lib="CDNA"
/clone="TCF-1a"
1. .1254
/gene="TCF-1"
/evidence=experimental
1. .1254
/gene="TCF-1"
CDS
80. .889
/gene="TCF-1"

```

```

/codon_start=1
/product="T cell factor 1, splice form A"
/protein_id="CAA42526.1"
/db_xref="GI:36786"
/db_xref="SWISS-PROT:P36402"
/translation="MYKTVYSAFNLLMHYPSPGAGOHPOPLHKANOPPHGVPO
LSLYEHFNSPHPTAPADISQKVHRPLOTDLSGFISLTSGSMGQLPHTVSWFTHPS
LMLGSGVPGHPAIPAIPAIYPPSGKOELOPDRNLKTOAESKAEEKKPTIKKPLNA
FMLYMKEAKRAVIAECTLKESAAINQILGRWHALSREFOAKYELARKERQLHMOLY
PGMSARDNYGKKRRSRKREHQSSTETTNMPRELNGQESLSMSSSSPA"
539. .769
misc_feature
/gene="TCF-1"
/note="HMG box"
810
misc_feature
/gene="TCF-1"
/note="alternative splice site"
BASE COUNT 320 a 391 c 335 g 208 t
ORIGIN

Query Match 94.1%; Score 16; DB 9; Length 1254;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccaccacc 16
|||||
Db 118 CATGCATTACCCACCC 133

RESULT 6
HSTCF1B HSTCF1B 2814 bp mRNA PRI 17-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form B).
DEFINITION X59870 X55329
ACCESSION X59870.1 GI:36787
VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES van de Wetering, M.
JOURNAL Direct Submission.
Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
REFERENCE 2 (bases 1 to 2814)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H.
TITLES Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box
JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 91114695
COMMENT See also X59869-X59871.
FEATURES
source
Location/Qualifiers
1. .2814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat and HPB-ALL"
/clone_lib="CDNA"
1. .2814
/gene="TCF-1"
/evidence=experimental
1. .2814
/gene="TCF-1"
CDS
80. .889
/gene="TCF-1"
/product="T cell factor 1 splice form B"
/protein_id="CAA42527.1"
/db_xref="GI:36788"
/db_xref="SWISS-PROT:P36402"
/translation="MYKTVYSAFNLLMHYPSPGAGOHPOPLHKANOPPHGVPO

```

LSLYEHFNSPHTTAPADISQKQVHRPLQTPDLSGFYSLTSGSMGOLPHTVSNFTHPS
 LMLGSGVCPHPAAIPHPAIVPPSGKQELQDPDRNLKTOAESKAEKAKPTIKKPLNA
 FMLYKMKRAKVIAECTLKESAAINOILGRRHALSREQAKYVELARKEROLHLMOLY
 PGWSARDNNGKKRSREKHQESITGGCKRNAFGTYPEKAAAPAPFLPMTVL"

misc_feature

539..769

/gene="TCF-1"

/note="HMG box"

misc_feature

810

/gene="TCF-1"

/note="alternative splice site"

763 a

831 c

648 g

572 t

Query Match 94.1%; Score 16; DB 9; Length 2814;

Best Local Similarity 100.0%; Pred. No. 5.5e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccacccc 16

|||||

Db 118 CATGCATTACCCACCC 133

RESULT 7

HSTCF1G

LOCUS HSTCF1G 2855 bp DNA PRI 30-SEP-1999

DEFINITION Homo sapiens TCF-1 gene.

ACCESSION X63901

VERSION X63901.1 GI:36791

KEYWORDS T-cell transcription factor; transcription factor.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2855)

van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D.,

Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.

The human T cell transcription factor-1 gene. Structure,

localization, and promoter characterization

J. Biol. Chem. 267 (12), 8530-8536 (1992)

92235082

2 (bases 1 to 2855)

Van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H.

Extensive alternative splicing and dual promoter usage generate

TCF-1 protein isoforms with differential transcription control

properties

Mol. Cell. Biol. 16 (3), 745-752 (1996)

96182076

3 (bases 1 to 2855)

Van de Wetering, M.L.

Direct Submission

Submitted (07-JAN-1992) M.L. Van De Wetering, Department of

Immunology, P.O. Box 8550, 3508 GA Utrecht, THE NETHERLANDS

Location/Qualifiers

1..2855

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome=5

/cell_type="T-lymphocyte"

/clone="phi-TCF-[5,8,11]"

/map="5q31.1"

829..924

/gene="TCF-1"

/number=1

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

/codon_start=1

/protein_id="CAB56795.1"

/db_xref="GI:6006565"

/translation="MYKTVYSAFNLLMHYPSPGACQHPQPPPLHKANQPHGVQ

LSLYEHFNSPHTTAPADISQKQVHRPLQTPDLSGFYSLTSGSMGOLPHTVSNFTHPS

LMLGSGVCPHPAAIPHPAIVPPSGKQELQDPDRNLKTOAESKAEKAKPTIKKPLNA

gene

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

925..970

/number=1

970..1074

/gene="TCF-1"

/number=2

1075..1119

/number=2

1120..1209

/gene="TCF-1"

/number=3

1210..1254

/number=3

1255..1374

/gene="TCF-1"

/number=4

1375..1419

/number=4

1420..1581

/gene="TCF-1"

/number=5

1582..1690

/number=5

1691..1798

/gene="TCF-1"

/number=6

1799..1842

/number=6

1843..1890

/gene="TCF-1"

/number=7

1891..2053

/number=1

2054..2134

/gene="TCF-1"

/number=8

592 a

929 c

771 g

562 t

1 others

BASE COUNT

ORIGIN

Query Match 94.1%; Score 16; DB 9; Length 2855;

Best Local Similarity 100.0%; Pred. No. 5.5e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccacccc 16

|||||

Db 867 CATGCATTACCCACCC 882

RESULT 8

HSTCF1G

LOCUS HSTCF1G 2910 bp mRNA PRI 17-JUN-1991

DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form C).

ACCESSION X59871 X55328

VERSION X59871.1 GI:36789

KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2910)

van de Wetering, M.

Direct Submission

Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical

Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

The Netherlands

2 (bases 1 to 2910)

van de Wetering, M., Oosterwegel, M., Dooyes, D. and Clevers, H.

Identification and cloning of TCF-1, a T lymphocyte-specific

transcription factor containing a sequence-specific HMG box
EMBO J. 10 (1), 123-132 (1991)

JOURNAL
MEDLINE
COMMENT

See also X59869-X59871.
Location/Qualifiers

source

1..2910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat and HPB-ALL"
/clone_lib="cDNA"

mRNA

1..2910
/gene="TCF-1"

gene

/evidence=experimental

CDS

1..2910
/gene="TCF-1"

80..886
/gene="TCF-1"

/codon_start=1

/product="T cell factor 1, splice form C"

/protein_id="CA42528.1"

/db_xref="GI:36790"

/translation="MYKTVYSAFNLLMHPYPPSGAGQHPOPPPLIKANOPPHGVQ
LSIYHEFNSPHTPAPADISQKVHRPLQTPDLGFSYLTSGSMGOLPHVSWFTFHS
LMLSGVPGHPAAIPHPATVPSPGKQELQPDRLKTOAESKAKEAKKPTIKKPLNA
FMLYGMEMRAKVIACETLKESAAINQILGRWHALSREEQAKYVELARKERQHMQLY
PGWSARDNYKKKRRSRKREKHQESTTDPGSPKKCRARFGLNQQTDMCGPCR"

misc_feature

539..769
/gene="TCF-1"

misc_feature

810
/note="HMG box"

/gene="TCF-1"

/note="alternative splice site"

BASE COUNT
ORIGIN

782 a 860 c 671 g 597 t

Query Match 94.1%; Score 16; DB 9; Length 2910;
Best Local Similarity 100.0%; Pred. No. 5.5e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccacc 16
|||||

Db 118 CATGCATTACCCACCC 133

RESULT 9

AC009012/c

LOCUS

DEFINITION

SEQUENCE, 7 ordered pieces.

AC009012

VERSION

AC009012.3 GI:12830104

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 84544)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 84544)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Feb 14, 2001 this sequence version replaced gi:7454202.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1189075, H71
Center clone name: XXpl-360D11

Summary Statistics

Consensus quality: 78189 bases at least Q40
Consensus quality: 82001 bases at least Q30
Consensus quality: 83201 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 83944; sum-of-contigs estimation
Quality coverage: 6.34 in Q20 bases; pulse field gel estimation
Quality coverage: 6.04 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 2979: contig of 2979 bp in length
* 2980 3079: gap of unknown length
* 3080 34576: contig of 31497 bp in length
* 34577 34676: gap of unknown length
* 34677 59262: contig of 24586 bp in length
* 59263 59362: gap of unknown length
* 59363 76010: contig of 16648 bp in length
* 76011 76110: gap of unknown length
* 76111 79613: contig of 3503 bp in length
* 79614 79713: gap of unknown length
* 79714 80649: contig of 936 bp in length
* 80650 80749: gap of unknown length
* 80750 84544: contig of 3795 bp in length.

FEATURES
source

1..84544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXpl-360D11"

BASE COUNT 19558 a 21740 c 22203 g 20376 t 667 others
ORIGIN

Query Match 94.1%; Score 16; DB 2; Length 84544;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccacc 16
|||||

Db 38701 CATGCATTACCCACCC 38686

RESULT 10

AC011336/c

LOCUS

DEFINITION

1 ordered pieces.

AC011336

VERSION

AC011336.4 GI:9256274

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 86914)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 86914)

DOE Joint Genome Institute.

Direct Submission

AC011336 86914 bp DNA HTG 23-APR-2001
Homo sapiens chromosome 5 clone CTC-250113, WORKING DRAFT SEQUENCE,
1 ordered pieces.

AC011336

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 86914)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 86914)

DOE Joint Genome Institute.

Direct Submission

JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 18, 2000 this sequence version replaced gi:7710593.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 281384, H247
Center clone name: CIT-HSPC_250113

Summary Statistics
Consensus quality: 82614 bases at least Q40
Consensus quality: 85602 bases at least Q30
Consensus quality: 86382 bases at least Q20
Estimated insert size: 89000; pulse field gel estimation
Estimated insert size: 86914; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 6.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 86914: contig of 86914 bp in length.
*
FEATURES
source
1..86914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-250113"
/clone_lib="Caltech human BAC library C"
BASE COUNT 20745 a 22135 c 22612 g 21422 t
ORIGIN
Query Match 94.1%; Score 16; DB 2; Length 86914:
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 catgcattaccaccc 16
|||||
Db 45276 CATGCATTACCACCC 45261

RESULT 11
AC009017/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone XXpl-929G6, WORKING DRAFT SEQUENCE,
26 unordered pieces.
ACCESSION AC009017
VERSION AC009017.4 GI:13699469
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 159420)
TITLE
Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159420)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:1178048.
-----Genome Center

Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1189133, H51
Center clone name: XXpl-929G6

Summary Statistics
Consensus quality: 142845 bases at least Q40
Consensus quality: 150910 bases at least Q30
Consensus quality: 152756 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 156920; sum-of-contigs estimation
Quality coverage: 8.41 in Q20 bases; pulse field gel estimation
Quality coverage: 4.29 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1423: contig of 1423 bp in length
* 1523: gap of unknown length
* 1524 3066: contig of 1543 bp in length
* 3067 3166: gap of unknown length
* 3167 5629: contig of 2463 bp in length
* 5630 5729: gap of unknown length
* 5730 6997: contig of 1268 bp in length
* 6998 7097: gap of unknown length
* 7098 9214: contig of 2117 bp in length
* 9215 9315: gap of unknown length
* 9316 11813: contig of 2499 bp in length
* 11814 11913: gap of unknown length
* 11914 14718: contig of 2805 bp in length
* 14719 14818: gap of unknown length
* 14819 17174: contig of 2356 bp in length
* 17175 17275: gap of unknown length
* 17276 19810: contig of 2436 bp in length
* 19811 21678: contig of 1868 bp in length
* 21679 21778: gap of unknown length
* 21779 26021: contig of 4243 bp in length
* 26022 26121: gap of unknown length
* 26122 29609: contig of 3488 bp in length
* 29610 29709: gap of unknown length
* 29710 33196: contig of 3487 bp in length
* 33197 33296: gap of unknown length
* 33297 36995: contig of 3699 bp in length
* 36996 37095: gap of unknown length
* 37096 40647: contig of 3551 bp in length
* 40648 40747: gap of unknown length
* 40748 45044: contig of 4298 bp in length
* 45045 45144: gap of unknown length
* 45145 52148: contig of 7004 bp in length
* 52149 52249: gap of unknown length
* 52250 57687: contig of 5438 bp in length
* 57688 66113: gap of unknown length
* 66114 66214: contig of 8327 bp in length
* 66215 75539: contig of 9325 bp in length
* 75540 75639: gap of unknown length
* 75640 82250: contig of 6611 bp in length
* 82251 82350: gap of unknown length
* 82351 95466: contig of 13116 bp in length
* 95467 95566: gap of unknown length
* 95567 110943: contig of 15377 bp in length
* 110944 127290: contig of 16248 bp in length
* 127291 143681: contig of 16291 bp in length
* 143682 143681: contig of 16291 bp in length

* 143682 143781: gap of unknown length
* 143782 159420: contig of 15639 bp in length.
FEATURES
source
1. .159420
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXPI-929G6"
BASE COUNT 39367 a 38414 c 39218 g 39915 t 2506 others
ORIGIN

Query Match 94.1%; Score 16; DB 2; Length 159420;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 catcattaccaccc 16
|||||
DB 56830 CATGCATTACCCACC 56815

RESULT 12
HS269M15 177562 bp DNA PRI 20-JUN-2001
LOCUS Human DNA sequence from clone RPI-269M15 on chromosome 20q12-13.12
DEFINITION Contains a gene similar to peptidylprolyl isomerase (cyclophilin),
and the 5' end of the PTPRT gene encoding protein tyrosine
phosphatase, receptor type, T, ESTs, STSs, GSSs and CpG Islands,
complete sequence.
ACCESSION AL021395
VERSION AL021395.16 GI:12406947
KEYWORDS HTG; CpG Island; PTPRT.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 177562)

COMMENT
Direct Submission
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Jan 23, 2001 this sequence version replaced gi:6249356.
requests: clonequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RPI-269M15 is from the library RPCI-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-269M15 The true
left end of clone RP5-112P14 is at 84910 in this sequence. The
true right end of clone RPI-151A11 is at 16712 in this sequence.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
Location/Qualifiers

source
1. .177562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q12-13.12"
/clone="RPI-269M15"
/clone_lib="RPCI-1"
15. .221
/note="MIR repeat: matches 26. .221 of consensus"
407. .593
/note="MIR repeat: matches 20. .214 of consensus"
1081. .1205
/note="I2 repeat: matches 2619. .2750 of consensus"
1407. .1593
/note="MER5A repeat: matches 1. .189 of consensus"
complement(1544..1966)
/note="match: GSS: Em:AQ880374"
1783. .1997
/note="LTR33 repeat: matches 312. .521 of consensus"
2215. .2334
/note="MER81 repeat: matches 1. .114 of consensus"
2408. .2641
/note="MIR repeat: matches 27. .258 of consensus"
2647. .2896
/note="AluSp repeat: matches 55. .303 of consensus"
complement(3322..3994)
/note="match: GSS: Em:AQ014397"
3711. .3813
/note="MIR repeat: matches 83. .190 of consensus"
3956. .4078
/note="MIR repeat: matches 3. .126 of consensus"
5269. .5319
/note="AluJ/FRAM repeat: matches 227. .278 of consensus"
5319. .5369
/note="AluJ/FRAM repeat: matches 227. .278 of consensus"
5369. .5419
/note="AluSc repeat: matches 232. .275 of consensus"
5410. .5469
/note="AluJ/FRAM repeat: matches 215. .278 of consensus"
5469. .5519
/note="AluJ/FRAM repeat: matches 227. .278 of consensus"
5518. .5569
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5568. .5619
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5618. .5669
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5668. .5719
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5718. .5769
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5768. .5819
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5818. .5869
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5868. .5919
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5918. .5969
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
5968. .6019
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
6018. .6069
/note="AluJ/FRAM repeat: matches 226. .278 of consensus"
6133. .6167
/note="Alu repeat: matches 76. .110 of consensus"
9619. .9669
/note="MT1E repeat: matches 1. .51 of consensus"
9677. .10127
/note="MLT2CA repeat: matches 1. .460 of consensus"
10130. .10181
/note="26 copies 2 mer ct 78% conserved"
10185. .10222
/note="19 copies 2 mer ag 78% conserved"

repeat_region	10248..10539 /note="AluSq repeat: matches 1..293 of consensus" 10602..10640 /note="MLR2CA repeat: matches 459..497 of consensus" 10629..11083 /note="MLR1E repeat: matches 9..504 of consensus" 11517..11716 /note="L1MC3 repeat: matches 6690..6901 of consensus" 12169..12469 /note="AluSx repeat: matches 1..299 of consensus" 12612..12641 /note="15 copies 2 mer aa 86% conserved" 13155..13285 /note="MIR repeat: matches 22..153 of consensus" 13311..13899 /note="L1MB8 repeat: matches 5536..6145 of consensus" 14184..14475 /note="AluSc repeat: matches 1..289 of consensus" 15433..15731 /note="AluSx repeat: matches 5..306 of consensus" 15860..16076 /note="MIR repeat: matches 20..241 of consensus" complement(16131..16704) /note="match: GSS: Em:AQ376253" 18613..20349 /note="CpG island" /evidence=not_experimental 19006..19030 /note="Single clone region. Assembly confirmed by restriction enzyme digest data." /gene="PTPRT" /note="match: cDNAs: Em:AF043644 Em:AF162857 Em:AF162856 Em:AF152556" /product="dJ269M15.2 (Protein tyrosine phosphatase, receptor type, T (RPTPRHO, KIAA0283))" /evidence=not_experimental complement(19423..19746) /gene="PTPRT" complement(<19423..19510) /gene="PTPRT" /note="Continues in dJ914M10 (Em:ALJ121763) The full gene is contained in the following clones Em:ALJ121763, Em:AL021395, Em:AL031656, Em:AL022239, Em:Z93942, Em:AL024473 and Em:AL035459, Em:AL049812, Em:AL136461 match: cDNAs: Em:AF043644 match: proteins: Tr:O43655" /codon_start=1 /evidence=not_experimental /product="dJ269M15.2 (Protein tyrosine phosphatase, receptor type, T (RPTPRHO, KIAA0283))" /protein_id="CAC24740.2" /db_xref="gi:14530721" /translation="MASLAALALLLRLLQLPLPGARQAQSA" 19684..19729 /note="23 copies 2 mer gg 73% conserved" 19747..19794 /note="16 copies 3 mer cgc 85% conserved" 20884..20941 /note="L1M4b repeat: matches -10..42 of consensus" 21254..21324 /note="L2 repeat: matches 2673..2742 of consensus" 21633..22115 /note="match: GSS: Em:AQ063598" 21894..22082 /note="L2 repeat: matches 2298..2490 of consensus" 22558..22605 /note="MIR repeat: matches 81..135 of consensus" 22715..22845 /note="L2 repeat: matches 2608..2747 of consensus" 23220..23359 /note="MER3 repeat: matches 1..152 of consensus"
repeat_region	23388..23477 /note="MER58A repeat: matches 29..118 of consensus" 23481..23505 /note="MER3 repeat: matches 1..25 of consensus"
Query Match	94.1%; Score 16; DB 9; Length 177562; Best Local Similarity 100.0%; Pred. No. 2e+02; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 catgcattaccaccc 16
Db	31134 CATGCATTACCCACCC 31149
RESULT	13
AC008608	200831 bp DNA HTG 20-APR-2001
LOCUS	Homo sapiens chromosome 5 clone CTB-113120, WORKING DRAFT SEQUENCE, 29 unordered pieces.
ACCESSION	AC008608
VERSION	AC008608.5 GI:13699394
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 200831) DOE Joint Genome Institute. Sequencing of Human Chromosome 5
TITLE	Unpublished
JOURNAL	2 (bases 1 to 200831) DOE Joint Genome Institute.
REFERENCE	Direct Submission Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Apr 20, 2001 this sequence version replaced gi:7708985.
TITLE	-----Genome Center
JOURNAL	Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov
COMMENT	----- Project Information Center Project Name: 107808, H233 Center clone name: CIT978SKB_113120 ----- Summary Statistics Consensus quality: 174733 bases at least Q40 Consensus quality: 186693 bases at least Q30 Consensus quality: 190547 bases at least Q20 Estimated insert size: 165000; pulse field gel estimation Estimated insert size: 198031; sum-of-contigs estimation Quality coverage: 7.77 in Q20 bases; pulse field gel estimation Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation. * NOTE: This is a 'working draft' sequence. It currently * consists of 29 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1115: contig of 1115 bp in length * 1116 1215: gap of unknown length * 1216 2567: contig of 1352 bp in length * 2568 2667: gap of unknown length * 2668 3745: contig of 1078 bp in length * 3746 3845: gap of unknown length * 3846 4921: contig of 1075 bp in length * 4921 5020: gap of unknown length * 5021 6045: contig of 1024 bp in length * 6045 6145: gap of unknown length * 6145 7285: contig of 1141 bp in length * 7286 7385: gap of unknown length

Thu Feb 21 07:25:52 2002

```
* 7386 8659: contig of 1274 bp in length
* 8660 8759: gap of unknown length
* 8760 8804: contig of 1045 bp in length
* 9804 9805: gap of unknown length
* 9905 11696: contig of 1792 bp in length
* 11697 11796: gap of unknown length
* 11797 12841: contig of 1045 bp in length
* 12842 12941: gap of unknown length
* 12942 15043: contig of 2102 bp in length
* 15044 15143: gap of unknown length
* 15144 17289: contig of 2146 bp in length
* 17290 17389: gap of unknown length
* 17390 20332: contig of 2843 bp in length
* 20333 21884: gap of unknown length
* 21885 24278: contig of 2294 bp in length
* 24279 24378: gap of unknown length
* 24379 28103: contig of 3725 bp in length
* 28104 28203: gap of unknown length
* 28204 30867: contig of 2664 bp in length
* 30868 30967: gap of unknown length
* 30968 34914: contig of 3947 bp in length
* 34915 35015: gap of unknown length
* 35016 36031: contig of 1016 bp in length
* 36032 36131: gap of unknown length
* 36132 38852: contig of 2722 bp in length
* 38853 38952: gap of unknown length
* 38953 41871: contig of 2919 bp in length
* 41872 41971: gap of unknown length
* 41972 47080: contig of 5109 bp in length
* 47081 47180: gap of unknown length
* 47181 49750: contig of 2570 bp in length
* 49751 49850: gap of unknown length
* 49851 57118: contig of 7268 bp in length
* 57119 57218: gap of unknown length
* 57219 60046: contig of 8828 bp in length
* 60047 66146: gap of unknown length
* 66147 81047: contig of 14900 bp in length
* 81048 81146: gap of unknown length
* 81147 101650: contig of 20504 bp in length
* 101651 101750: gap of unknown length
* 101751 137108: contig of 35358 bp in length
* 137109 137208: gap of unknown length
* 137209 200831: contig of 63623 bp in length.
FEATURES             Location/Qualifiers
     source           1..200831
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone_lib="Caltech human BAC library B"
BASE COUNT           53266 a 46907 c 47140 g 50695 t 2823 others
ORIGIN

Query Match          94.1%; Score 16; DB 2; Length 200831;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcattaccaccc 16
Db 184024 CATGCATTACCCACCC 184039

RESULT 14
LOCUS      HSPF15G5      216 bp      DNA      STS      21-MAY-1998
DEFINITION H.sapiens flow-sorted chromosome 20 HindIII fragment., SC20PF15G5,
sequence tagged site.
ACCESSION 294649
VERSION   294649.1 GI:1946134
KEYWORDS  STS; single read.

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 216)
AUTHORS    Deloukas,P., Buck,D., Langford,C., Ross,M.T. and Hunt,S.E.
TITLE      Direct Submission
JOURNAL    Submitted (17-APR-1997) The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
humquerry@sanger.ac.uk
COMMENT     Vector: pBSIIISK+
Marker STG25573 (Primer A : TCATGCATTACCCACCAAGA; Primer B :
AAGGAATGGAAGAGAGAGAAGG; amplicon size : 145 bp) was mapped to
chromosome 20 using Radiation Hybrid panel Genebridge 4 (G94).
FEATURES             Location/Qualifiers
     source           1..216
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /sex="Female"
                     /dev_stage="adult"
                     /tissue_type="EBV lymphoblastoid cell line"
                     /clone_lib="SC20PF"
                     /clone="SC20PF15G5"
BASE COUNT           58 a 44 c 17 g 87 t
ORIGIN

Query Match          90.6%; Score 15.4; DB 11; Length 216;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccacccca 17
Db 60 CATGCATTACCCACCAA 76

RESULT 15
LOCUS      AF097994      2046 bp      mRNA      PRI      01-NOV-1999
DEFINITION Homo sapiens L-kynurenine/alpha-aminoacidate aminotransferase mRNA,
complete cds.
ACCESSION  AF097994
VERSION     AF097994.1 GI:6165633
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 2046)
AUTHORS    Gattai,S., Breton,J., Mostardini,M., Mosca,M., Tarroni,P.,
Schwarz,R., Speciale,C., Okuno,E., Toma,S. and Benatti,L.
TITLE      Direct Submission
JOURNAL    Submitted (09-OCT-1998) R&D - CNS, Pharmacia and Upjohn, V. le
Pasteur, 10, Nerviano, Milano 20014, Italy
FEATURES             Location/Qualifiers
     source           1..2046
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /codon_start=1
                     /product="L-kynurenine/alpha-aminoacidate
aminotransferase"
                     /protein_id="AAF04623.1"
                     /db_xref="GI:6165634"
                     /translation="MNYARFITAASARNPSPIRTMTDILSHCKPSMISLAGLNPNP
MPPPKTAVITVENGCKTQFGEMMKRALQYSFSGAGIPELLISWLKQLKLNPPPTIHY
CDS
ORIGIN
```


PPSQGMDCVTSQSGQGLCKVFEIINPGDNVLLDEPAYSGTQLSHPLGCNIIINVA
SDESGIVPDSLRLDILSRWKPEDAKNPOTPKFLYTVPNGNNPTGNSLTSEKKKEIYE
LARKYDFLIIEDDPYFLOFNKFRVPTFLSMDVDGRVIRADFSKIISSGLRIGFLTG
PKPLIERVILHIQVSTLHPSTFNQLMISOLLHEWGEFMAHVDRVIDFYSNOKDAIL
AAADRWLTLAEWHVPAAGMFLIKYKGINDVKELIEEKAVKMGVLMPLPCNRFYVDSS
APSPYLRAFSFSASPEQMDVAFQVLAQLIKESL*

BASE COUNT 608 a 457 c 467 g 514 t
ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 2046;
Best Local Similarity 94.1%; Pred. NO. 1.2e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 catgcattaccaccca 17
||| |||||
Db 750 CATCCATTACCCACCCA 766

Search completed: February 20, 2002, 11:01:42
Job time: 12010 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:41:32 ; Search time 363.89 Seconds
(without alignments)
40.052 Million cell updates/sec

Title: US-09-904-420A-4

Perfect score: 17
Sequence: 1 catgattaccaccacca 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.4	90.6	243	AAZ46032	Partial cDNA encod
2	15.4	90.6	590	AA139761	Probe #8447 used t
3	15.4	90.6	2125	AAZ46023	cDNA encoding a hu
4	15.4	90.6	2348	AA159213	Human polynucleoti
5	15	88.2	2184	AAH18256	Human cDNA sequenc
6	15	88.2	2737	AA159251	Human polynucleoti
7	14.4	84.7	289	AAH83386	Breast cancer tumo
8	14.4	84.7	289	AAV68924	DNA molecule encod
9	14.4	84.7	289	AAH80909	Human breast tumou
10	14.4	84.7	378	AAH29229	Drosophila melanog
11	14.4	84.7	401	AAV66258	HGV sequence ampli

C 12	14.4	84.7	401	19	AAV66259	HGV sequence ampli
C 13	14.4	84.7	401	19	AAV66260	HGV sequence ampli
C 14	14.4	84.7	401	19	AAV66261	HGV variant 7244 D
C 15	14.4	84.7	401	19	AAV66222	HGV variant K27 DN
C 16	14.4	84.7	401	19	AAV66223	HGV variant K30 DN
C 17	14.4	84.7	401	20	AAV16533	Polynucleotide seq
C 18	14.4	84.7	401	20	AAV16534	Polynucleotide seq
C 19	14.4	84.7	401	20	AAV16535	Polynucleotide seq
C 20	14.4	84.7	401	20	AAV62496	US5856134 Seq ID 2
C 21	14.4	84.7	401	20	AAV62497	US5856134 Seq ID 2
C 22	14.4	84.7	401	20	AAV62498	US5856134 Seq ID 2
C 23	14.4	84.7	401	20	AAV82276	Hepatitis G virus
C 24	14.4	84.7	401	20	AAV82277	Hepatitis G virus
C 25	14.4	84.7	401	20	AAV82278	Hepatitis G virus
C 26	14.4	84.7	402	19	AAV66247	HGV sequence ampli
C 27	14.4	84.7	402	19	AAV66248	HGV variant 156957
C 28	14.4	84.7	402	20	AAV16522	Polynucleotide seq
C 29	14.4	84.7	402	20	AAV62485	US5856134 Seq ID 2
C 30	14.4	84.7	402	20	AAV82265	Hepatitis G virus
C 31	14.4	84.7	420	19	AAV63636	Schizochytrium sp.
C 32	14.4	84.7	420	20	AAV00914	M. alpina desatura
C 33	14.4	84.7	420	20	AAV82635	Partial DNA sequen
C 34	14.4	84.7	420	21	AAA09446	Schizochytrium sp.
C 35	14.4	84.7	420	21	AAA14587	DNA encoding a par
C 36	14.4	84.7	436	18	AAV96420	Hepatitis GB virus
C 37	14.4	84.7	632	18	AAV96405	Hepatitis GB virus
C 38	14.4	84.7	2484	20	AAZ11192	Human cerebellin-2
C 39	14.4	84.7	2751	22	AAZ11286	Human cDNA sequenc
C 40	14.4	84.7	8630	19	AAV23082	Fragment HGV1775 o
C 41	14.4	84.7	9256	19	AAV23075	Fragment HGV606 of
C 42	14.4	84.7	9271	19	AAV23080	Fragment HGV2141 o
C 43	14.4	84.7	9331	19	AAV23079	Fragment HGV1916 o
C 44	14.4	84.7	9333	19	AAV23084	Fragment k1029 of
C 45	14.4	84.7	9390	19	AAV23078	Fragment HGV1789 o

ALIGNMENTS

RESULT 1

AAZ46032
ID AAZ46032 standard; cDNA; 243 BP.

XX AAZ46032:

AC
XX
DT 25-APR-2000 (first entry)

DE Partial cDNA encoding a transferase designated HUTRAN-2 fragment.

XX Transferase; HUTRAN-2; kynurenine/alpha-aminoadipate aminotransferase;
KW autoimmune disorder; inflammatory disorder; AIDS; atherosclerosis;
KW adult respiratory distress syndrome; allergy; asthma; trauma;
KW autoimmune thyroiditis; bronchitis; Crohn's disease; diabetes mellitus;
KW gout; Grave's disease; osteoarthritis; osteoporosis; pancreatitis;
KW psoriasis; rheumatoid arthritis; infection; neurological disorder;
KW epilepsy; ischemic cerebrovascular disease; stroke; Alzheimer's disease;
KW Pick's disease; Huntington's disease; dementia; Parkinson's disease;
KW extrapyramidal disorder; viral central nervous system disease;
KW prion disease; central nervous system developmental disorder;
KW neuroskeletal disorder; muscular dystrophy; neuromuscular disorder;
KW peripheral nervous system disorder; mental disorder; schizophrenia;
KW anxiety; reproductive disorder; gastrointestinal disorder; dyspepsia;
KW indigestion; gastritis; anorexia; nausea; abdominal angina;
KW gastroenteritis; intestinal obstruction; peptic ulcer;
KW irritable bowel syndrome; diarrhoea; constipation;
KW gastrointestinal haemorrhage; cancer; ss.

OS Homo sapiens.

XX WO200000594-A2.

PN

XX 06-JAN-2000.

PD

XX

Thu Feb 21 07:25:54 2002

us-09-904-420a-4.ing

```

PF 29-JUN-1999; 99WO-US14651.
XX
PR 30-JUN-1999; 98US-0109204.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P. Bandman O, Hillman JL, Guegler KJ, Gorgone GA, Corley NC;
XX Patterson C;
XX WPI: 2000-147267/13.
XX
XX Novel human transferases used for the diagnosis, treatment, and
XX prevention of autoimmune/inflammatory, neurological, reproductive and
XX gastrointestinal disorders and cancer
XX
XX Disclosure; Page 86; 95pp; English.
XX
XX AAZ46032-38 represent partial fragments which were used to produce
XX Incyte clone 2150892, which encodes a transferase polypeptide,
XX designated HUTRAN-2. The polypeptide is a kynurenine/alpha-aminoadipate
XX aminotransferase. The HUTRAN polypeptides, polynucleotides, agonists,
XX antagonists, and antibodies can be used to diagnose, treat or prevent
XX autoimmune/inflammatory diseases (e.g. AIDS, adult respiratory distress
XX syndrome, allergies, asthma, atherosclerosis, autoimmune thyroiditis,
XX bronchitis, Crohn's disease, diabetes mellitus, gout, Grave's disease,
XX osteoarthritis, osteoporosis, pancreatitis, psoriasis, rheumatoid
XX arthritis, infections, trauma, neurological disorders (e.g. epilepsy,
XX ischemic cerebrovascular disease, stroke, Alzheimer's disease, Picks
XX disease, Huntington's disease, dementia, Parkinson's disease and other
XX extrapyramidal disorders, viral central nervous system disease, prion
XX diseases, central nervous system developmental disorders, neuroskeletal
XX disorders, muscular dystrophy, neuromuscular disorders, peripheral
XX nervous system disorders, mental disorders, schizophrenic, anxiety,
XX reproductive disorders and gastrointestinal disorders (e.g. dyspepsia,
XX indigestion, gastritis, anorexia, nausea, abdominal angina,
XX gastroenteritis, intestinal obstruction, intestinal tract infection,
XX peptic ulcer, irritable bowel syndrome, diarrhoea, constipation,
XX gastrointestinal haemorrhage, and cancer.
XX
XX Sequence 243 BP; 75 A; 54 C; 48 G; 65 T; 1 other;

Query Match 90.6%; Score 15.4; DB 21; Length 243;
Best Local Similarity 94.1%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17
Db 121 catccattaccaccacca 137

RESULT 2
AAI39761/C
ID AAI39761 standard; DNA; 590 BP.
XX
XX AAI39761;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #8447 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX

26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI: 2001-488897/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human placenta
Claim 25; SEQ ID No 8447; 654pp; English.
The present invention relates to single exon nucleic acid probes (SENP).
The present sequence is one such probe. The probes are useful for
producing a microarray for predicting, measuring and displaying gene
expression in samples derived from human placenta. The probes are useful
for antenatal diagnosis of human genetic disorders.
Sequence 590 BP; 198 A; 99 C; 149 G; 144 T; 0 other;

Query Match 90.6%; Score 15.4; DB 22; Length 590;
Best Local Similarity 94.1%; Pred. No. 42;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17
Db 24 CATGCTTTACCCACCCA 8

RESULT 3
AAZ46023
TD AAZ46023 standard; cDNA; 2125 BP.
XX
XX AAZ46023;
XX
XX 25-APR-2000 (first entry)
XX
XX cDNA encoding a human transferase designated HUTRAN-2.
XX
XX Transferase: HUTRAN-2; kynurenine/alpha-aminoadipate aminotransferase;
XX autoimmune disorder; inflammatory disorder; AIDS; atherosclerosis;
XX adult respiratory distress syndrome; allergy; asthma; trauma;
XX autoimmune thyroiditis; bronchitis; Crohn's disease; diabetes mellitus;
XX gout; Grave's disease; osteoarthritis; osteoporosis; pancreatitis;
XX psoriasis; rheumatoid arthritis; infection; neurological disorder;
XX epilepsy; ischemic cerebrovascular disease; stroke; Alzheimer's disease;
XX Picks disease; Huntington's disease; dementia; Parkinson's disease;
XX extrapyramidal disorder; viral central nervous system disease;
XX prion disease; central nervous system developmental disorder;
XX neuroskeletal disorder; muscular dystrophy; neuromuscular disorder;
XX peripheral nervous system disorder; mental disorder; schizophrenic;
XX anxiety; reproductive disorder; gastrointestinal disorder; dyspepsia;
XX indigestion; gastritis; anorexia; nausea; abdominal angina;
XX gastroenteritis; intestinal obstruction; peptic ulcer;
XX irritable bowel syndrome; diarrhoea; constipation;
XX gastrointestinal haemorrhage; cancer; ss.
XX
XX Homo sapiens.
XX
XX OS
XX Key Location/Qualifiers
XX CDS 150..1427
XX /*tag= a
XX /product= "transferase"
XX
XX WO200000594-A2.
XX

```

Thu Feb 21 07:25:54 2002

us-09-904-420a-4.rng

```
PD 06-JAN-2000.
XX
XX
XX 29-JUN-1999; 99WO-US14651.
XX
XX 30-JUN-1998; 98US-0109204.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Bandman O, Hillman JL, Guegler KJ, Gorgone GA, Corley NC;
XX Patterson C;
XX
XX WPI: 2000-147267/13.
XX P-PSDB; AAY54592.
XX
XX Novel human transferases used for the diagnosis, treatment, and
XX prevention of autoimmune/inflammatory, neurological, reproductive and
XX gastrointestinal disorders and cancer -
XX
XX Claim 7; Page 81; 95pp; English.
XX
XX The present sequence encodes a transferase polypeptide, designated
XX HUTRAN-2. The HUTRAN cDNA sequence is derived from Incyte clone 2150892.
XX The sequence is obtained from the partial sequences AA246032-38. The
XX polypeptide is a kynurenine/alpha-aminoacidate aminotransferase. The
XX HUTRAN polypeptides, polynucleotides, agonists, antagonists, and
XX antibodies can be used to diagnose, treat or prevent autoimmune/
XX inflammatory diseases (e.g. AIDS, adult respiratory distress syndrome,
XX allergies, asthma, atherosclerosis, autoimmune thyroiditis, bronchitis,
XX Crohn's disease, diabetes mellitus, gout, Grave's disease,
XX osteoarthritis, osteoporosis, pancreatitis, psoriasis, rheumatoid
XX arthritis, infections, trauma, neurological disorders (e.g. epilepsy,
XX ischemic cerebrovascular disease, stroke, Alzheimer's disease, Picks
XX disease, Huntington's disease, dementia, parkinson's disease and other
XX extrapyramidal disorders, viral central nervous system disease, prion
XX diseases, central nervous system developmental disorders, neuroskeletal
XX disorders, muscular dystrophy, neuromuscular disorders, peripheral
XX nervous system disorders, mental disorders, schizophrenic, anxiety,
XX reproductive disorders and gastrointestinal disorders (e.g. dyspepsia,
XX indigestion, gastritis, anorexia, nausea, abdominal angina,
XX gastroenteritis, intestinal obstruction, intestinal tract infections,
XX peptic ulcer, irritable bowel syndrome, diarrhoea, constipation,
XX gastrointestinal haemorrhage, and cancer.
XX
XX Sequence 2125 BP; 563 A; 424 C; 428 G; 610 T; 0 other;
XX
XX
XX Query Match 90.6%; Score 15.4; DB 21; Length 2125;
XX Best Local Similarity 94.1%; Pred. No. 46;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 catgcattaccaccacca 17
XX ||| ||||| ||||| |||||
XX Db 446 catccattaccaccacca 462
XX
XX RESULT 4
XX ID AA159213
XX AC AA159213 standard; cDNA: 2348 BP.
XX
XX XX
XX XX 22-OCT-2001 (first entry)
XX
XX XX Human polynucleotide SEQ ID NO 1416.
XX
XX XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemolactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX XX Homo sapiens.
XX
XX OS
```

```
XX
XX PN WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 14-AUG-2000; 2000US-0653450.
XX 13-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB; AAM40057.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1416; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAH42213) with nototropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX
XX Note: the sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 2348 BP; 722 A; 488 C; 492 G; 646 T; 0 other;
XX
XX
XX Query Match 90.6%; Score 15.4; DB 22; Length 2348;
XX Best Local Similarity 94.1%; Pred. No. 46;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 catgcattaccaccacca 17
XX ||| ||||| ||||| |||||
XX Db 650 catccattaccaccacca 666
XX
XX RESULT 5
XX AAH18256
XX ID AAH18256 standard; cDNA: 2184 BP.
XX
XX XX
XX XX AAH18256;
XX
XX XX 26-JUN-2001 (first entry)
XX
XX XX Human cDNA sequence SEQ ID NO:18214.
XX DE
XX XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX KW
XX XX Homo sapiens.
XX
XX OS
```

PN EP1074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 DR Primer sets for synthesizing polynucleotides, particularly the 5602
 XX full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PT
 XX Claim 8; SEQ ID 18214; 2537pp + CD ROM; English.
 PS The present invention describes primer sets for synthesising 5602
 XX full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 2184 BP; 458 A; 635 C; 635 G; 456 T; 0 other;
 SQ
 Query Match 88.2%; Score 15; DB 22; Length 2184;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 atgcattaccaccc 16
 DB 1337 atgcattaccaccc 1351
 RESULT 6
 ID AA159251 standard; cDNA; 2737 BP.
 XX AA159251;
 AC AA159251;
 XX 22-OCT-2001 (first entry)
 DT Human polynucleotide SEQ ID NO 1454.
 DE Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX Homo sapiens.
 OS WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AAM40095.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 XX Claim 1; SEQ ID NO 1454; 10078pp; English.
 PS The invention relates to human nucleic acids (AA157798-AA161369) and
 XX the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 2737 BP; 484 A; 840 C; 899 G; 514 T; 0 other;
 SQ
 Query Match 88.2%; Score 15; DB 22; Length 2737;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 atgcattaccaccc 16
 DB 2622 atgcattaccaccc 2636
 RESULT 7
 AAH83386/c
 ID AAX83386 standard; cDNA; 289 BP.
 XX AAX83386;
 AC AAX83386;
 XX 31-AUG-1999 (first entry)
 DT
 XX

DE Breast cancer tumour specific clone #140.
 XX
 KW Breast cancer; tumour; gene expression; genome; diagnosis; mammary;
 KW human endogenous retrovirus; vaccine; ss.
 XX
 OS Homo sapiens.
 XX W09725426-A2.
 PN
 XX
 PD 17-JUL-1997.
 XX
 XX 10-JAN-1997; 97WO-US00485.
 XX
 PR 20-AUG-1996; 96US-0700014.
 PR 11-JAN-1996; 96US-0585392.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 PI Frudakis TN, Reed SG, Smith JM;
 XX WPI: 1997-372865/34.
 DR
 XX Breast cancer-related DNA from retrovirus antigen (s) - useful for
 PT diagnosis and treatment of breast cancer
 PT
 XX Claim 1: Page 167; 221pp; English.
 XX
 CC Sequences AAX83201-X83285 and AAX83331-X83415 represent novel breast
 CC cancer tumour specific clones which are expressed from a genomic region
 CC containing a human endogenous retrovirus (AAX83330). Detection of
 CC the clone sequences allows determination of the presence of breast
 CC cancer in a mammal. Progression of breast cancer can be monitored by
 CC detecting the level of clone expression. Polypeptides encoded by the
 CC clones can be used in vaccines to inhibit or prevent breast cancer.
 XX
 XX Sequence 289 BP; 101 A; 48 C; 67 G; 73 T; 0 other;
 SQ

Query Match 84.7%; Score 14.4; DB 18; Length 289;
 Best Local Similarity 93.8%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 atgcattaccaccacca 17
 II IIIIIIIIIIIII
 Db 77 ATACATTACCCACCCA 62

RESULT 8
 AAV68924/C
 ID AAV68924 standard; DNA; 289 BP.
 XX
 AC AAV68924;
 XX
 XX 22-JAN-1999 (first entry)
 XX
 XX DNA molecule encoding a breast tumour specific polypeptide #116.
 XX
 DE Human; breast cancer; breast tumour tissue; diagnosis; treatment;
 KW vaccine; epitope; endogenous; retroviral element; ss.
 KW
 XX Homo sapiens.
 OS
 XX W09845328-A2.
 PN
 XX
 PD 15-OCT-1998.
 XX
 XX 09-APR-1998; 98WO-US06939.
 PF
 XX 11-DEC-1997; 97US-0991789.
 PR
 PR 09-APR-1997; 97US-0838762.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 PI Frudakis TN, Reed SG, Smith JM, Reed SG, Misher LE, Retter MW, Dillon DC;
 XX WPI: 2000-628403/60.
 DR
 XX An isolated polypeptide comprising an immunogenic portion of a breast
 PT tumor protein used for inhibiting the development of cancer, especially
 PT breast cancer, and monitoring cancer progression in a patient -
 XX
 XX Claim 4; Page 141; 187pp; English.
 PS
 XX The present sequence is given in a specification relating to compositions
 CC and methods for the treatment and diagnosis of breast cancer. Nucleotide
 CC sequences that are preferentially expressed in breast tumour tissue, and
 CC the polypeptides encoded by such nucleotide sequences, are used in

PI Frudakis TN, Reed SG, Smith JM;
 XX WPI: 1998-557473/47.
 XX
 PT New DNA sequences isolated from endogenous human retroviral element
 PT - and related vectors, transformed cells, proteins and antibodies,
 PT useful for diagnosis, treatment and prevention of breast cancer
 XX
 XX Claim 1; Page 101; 173pp; English.
 PS
 XX AAV68800 to AAV68998 represent nucleotide sequences which encode human
 CC breast tumour specific polypeptides. Detection or measurement of
 CC human breast tumour specific polypeptides and nucleotide sequences,
 CC or the corresponding RNA in a sample, is used for diagnosis and
 CC monitoring of breast cancer. Human breast tumour specific polypeptides
 CC and nucleotide sequences, and the vectors containing the DNAs, are also
 CC useful in vaccines for inhibiting development (for prevention or
 CC therapy) of breast cancer. The polypeptides may also be used to
 CC raise monoclonal antibodies, used as immunoassay reagents.
 XX
 SQ Sequence 289 BP; 101 A; 48 C; 67 G; 73 T; 0 other;

Query Match 84.7%; Score 14.4; DB 19; Length 289;
 Best Local Similarity 93.8%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 atgcattaccaccacca 17
 II IIIIIIIIIIIII
 Db 77 ATACATTACCCACCCA 62

RESULT 9
 AAC80309/C
 ID AAC80909 standard; cDNA; 289 BP.
 XX
 AC AAC80909;
 XX
 XX 13-FEB-2001 (first entry)
 DT
 XX Human breast tumour-specific cDNA SEQ ID NO: 197.
 DE
 XX Human; breast tumour-specific antigen; cytostatic; vaccine;
 KW breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.
 KW
 XX Homo sapiens.
 OS
 XX W0200061753-A2.
 PN
 XX 19-OCT-2000.
 PD
 XX 07-APR-2000; 2000WO-US09312.
 PF
 XX 09-APR-1999; 99US-0289198.
 PR
 PR 28-OCT-1999; 99US-0429755.
 PR
 PR 23-MAR-2000; 2000US-0534825.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 PI Frudakis TN, Smith JM, Reed SG, Misher LE, Retter MW, Dillon DC;
 XX WPI: 2000-628403/60.
 DR
 XX An isolated polypeptide comprising an immunogenic portion of a breast
 PT tumor protein used for inhibiting the development of cancer, especially
 PT breast cancer, and monitoring cancer progression in a patient -
 XX
 XX Claim 4; Page 141; 187pp; English.
 PS
 XX The present sequence is given in a specification relating to compositions
 CC and methods for the treatment and diagnosis of breast cancer. Nucleotide
 CC sequences that are preferentially expressed in breast tumour tissue, and
 CC the polypeptides encoded by such nucleotide sequences, are used in

CC compositions and vaccines to inhibit the development of cancer,
 CC especially breast cancer. The progression of a cancer may be monitored by
 CC carrying out detection of tumour-specific antigens at subsequent time
 CC points and comparing the results from the different time points.
 CC CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
 CC with tumour-specific polypeptides, polynucleotides encoding the
 CC polypeptides or antigen presenting cells expressing the polypeptides. The
 CC cells are then administered to the patient to inhibit development of
 CC cancer.
 XX
 SQ Sequence 289 BP; 101 A; 48 C; 67 G; 73 T; 0 other;

Query Match 84.7%; Score 14.4; DB 21; Length 289;
 Best Local Similarity 93.8%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcattaccaccaca 17
 |||||
 Db 77 ATACATTACCCACCA 62

RESULT 10

AAH29229
 ID AAH29229 standard; DNA; 378 BP.

XX
 AC AAH29229;

XX
 DT 17-JUL-2001 (first entry)

XX Drosophila melanogaster essential gene fragment, SEQ ID NO: 418.

XX Drosophila melanogaster; fruit fly; essential gene; screening assay;
 KW pesticide; crop protection; chromosome 3; ds.

XX Drosophila melanogaster.

XX WO200118547-A1.

XX 15-MAR-2001.

XX 06-SEP-2000; 2000WO-GB03444.

XX 07-SEP-1999; 99GB-0021009.

XX (UNIU) UNIV GLASGOW.

XX Davies RW, Kaiser K, Yang MY;

XX WPI; 2001-281436/29.

XX Screening assays for used for identifying compounds having a
 PT physiological effect on proteins identified as being essential -

XX Claim 1; Page 417; 695pp; English.

XX The present sequence is part of an essential gene from *Drosophila*
 CC melanogaster. Lack of expression of the protein encoded by this
 CC gene leads to a lethal or semi-lethal phenotype. The invention
 CC relates to 902 nucleic acid sequences from genes encoding proteins
 CC which are thought to be essential, and to a screening assay for
 CC identifying compounds which have a physiological effect on these
 CC proteins. Suitable compounds are useful as pesticides and may be used
 CC in conjunction with other pesticides and herbicides for crop
 CC protection. The gene corresponding to the present sequence is located
 CC on chromosome 3.

XX Sequence 378 BP; 122 A; 77 C; 86 G; 93 T; 0 other;

Query Match 84.7%; Score 14.4; DB 22; Length 378;
 Best Local Similarity 93.8%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcattaccaccaca 17
 |||||
 Db 135 atgcattaccaccaca 150

RESULT 11

AAV66258/C
 ID AAV66258 standard; DNA; 401 BP.

XX
 AC AAV66258;

XX 23-DEC-1998 (first entry)

XX HGV sequence amplified from serum code 7244.

XX PNF2161 sera; antigen; immunoreaction screening;

KW Non-A Non-B Non-C Non-D Non-E Hepatitis virus; HGV;

KW anti-HGV antibody; vaccine; ds.

XX Hepatitis G virus.

XX US5824507-A.

XX 20-OCT-1998.

XX 19-MAY-1995; 95US-0444733.

XX 19-MAY-1995; 95US-0444733.

XX 20-MAY-1994; 94US-0246985.

XX 03-AUG-1994; 94US-0285543.

XX 03-AUG-1994; 94US-0285561.

XX 26-OCT-1994; 94US-0329729.

XX 23-NOV-1994; 94US-0344271.

XX 16-DEC-1994; 94US-0357509.

XX 15-FEB-1995; 95US-0389886.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Fry KE, Kim JP, Linnen JM, Wages J, Young LM;

XX WPI; 1998-582553/49.

XX Polypeptide antigens hepatitis G virus - useful as vaccines against
 PT the virus and raising antibodies for use in immuno-testing for the
 PT virus

XX Example 18; Columns 323-324; 206pp; English.

XX AAV66231-66 represent the Hepatitis G virus (HGV) sequences, amplified
 CC from different sera by the PCR primers AAV66181-82). The specification
 CC describes HGV antigen clones isolated from PNF2161 sera. This sera
 CC is obtained from a patient suffering from Non-A Non-B Non-C
 CC Non-D Non-E Hepatitis Virus, also known as Hepatitis G virus (HGV).
 CC HGV antigens can be used in immunoreaction screening for presence of
 CC anti-HGV specific antibodies, especially in patient sera, e.g. enzyme
 CC linked immunoassay (ELISA). They can also be used as vaccines, and to
 CC raise the polyclonal and monoclonal antibodies.

XX Sequence 401 BP; 76 A; 112 C; 137 G; 76 T; 0 other;

Query Match 84.7%; Score 14.4; DB 19; Length 401;
 Best Local Similarity 93.8%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccacc 16
 |||||
 Db 393 CATGCATTCCACCC 378

RESULT 12
 AAV66259/c

ID AAV66259 standard; DNA; 401 BP.
 XX AC AAV66259;
 XX 23-DEC-1998 (first entry)
 DT XX HGV sequence amplified from serum code K27.
 DE XX
 XX PNF2161 sera; antigen; immunoreaction screening;
 KW Non-A Non-B Non-C Non-D Non-E Hepatitis virus; HGV;
 KW anti-HGV antibody; vaccine; ds.
 XX Hepatitis G virus.
 OS US5824507-A.
 XX 20-OCT-1998.
 XX 19-MAY-1995; 95US-0444733.
 XX 19-MAY-1995; 95US-0444733.
 PR 20-MAY-1994; 94US-0246985.
 PR 03-AUG-1994; 94US-0285543.
 PR 03-AUG-1994; 94US-0285561.
 PR 26-OCT-1994; 94US-0329729.
 PR 23-NOV-1994; 94US-0344271.
 PR 16-DEC-1994; 94US-0357509.
 PR 15-FEB-1995; 95US-0389886.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 PA Fry KE, Kim JP, Linnen JM, Wages J, Young LM;
 PI WPI; 1998-582553/49.
 XX Polypeptide antigens hepatitis G virus - useful as vaccines against
 PT the virus and raising antibodies for use in immuno-testing for the
 PT virus
 XX Example 18; Columns 323-326; 206pp; English.
 PS AAV66231-66 represent the Hepatitis G virus (HGV) sequences, amplified
 CC from different sera by the PCR primers AAV66181-82). The specification
 CC describes HGV antigen clones isolated from PNF2161 sera. This sera
 CC is obtained from a patient suffering from Non-A Non-B Non-C
 CC Non-D Non-E Hepatitis Virus, also known as Hepatitis G virus (HGV).
 CC HGV antigens can be used in immunoreaction screening for presence of
 CC anti-HGV specific antibodies, especially in patient sera, e.g. enzyme
 CC linked immunoassay (ELISA). They can also be used as vaccines, and to
 CC raise the polyclonal and monoclonal antibodies.
 XX Sequence 401 BP; 77 A; 112 C; 136 G; 76 T; 0 other;
 SQ

Query Match 84.7%; Score 14.4; DB 19; Length 401;
 Best Local Similarity 93.8%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccaccc 16
 DB 393 CATGCATTCCACCC 378

RESULT 13
 ID AAV66260 standard; DNA; 401 BP.
 XX AC AAV66260;
 XX 23-DEC-1998 (first entry)
 DT XX HGV sequence amplified from serum code K30.
 DE XX

KW PNF2161 sera; antigen; immunoreaction screening;
 KW Non-A Non-B Non-C Non-D Non-E Hepatitis virus; HGV;
 KW anti-HGV antibody; vaccine; ds.
 XX Hepatitis G virus.
 OS US5824507-A.
 XX 20-OCT-1998.
 XX 19-MAY-1995; 95US-0444733.
 XX 19-MAY-1995; 95US-0444733.
 PR 20-MAY-1994; 94US-0246985.
 PR 03-AUG-1994; 94US-0285543.
 PR 03-AUG-1994; 94US-0285561.
 PR 26-OCT-1994; 94US-0329729.
 PR 23-NOV-1994; 94US-0344271.
 PR 16-DEC-1994; 94US-0357509.
 PR 15-FEB-1995; 95US-0389886.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 PA Fry KE, Kim JP, Linnen JM, Wages J, Young LM;
 PI WPI; 1998-582553/49.
 XX Polypeptide antigens hepatitis G virus - useful as vaccines against
 PT the virus and raising antibodies for use in immuno-testing for the
 PT virus
 XX Example 18; Columns 325-326; 206pp; English.
 PS AAV66231-66 represent the Hepatitis G virus (HGV) sequences, amplified
 CC from different sera by the PCR primers AAV66181-82). The specification
 CC describes HGV antigen clones isolated from PNF2161 sera. This sera
 CC is obtained from a patient suffering from Non-A Non-B Non-C
 CC Non-D Non-E Hepatitis Virus, also known as Hepatitis G virus (HGV).
 CC HGV antigens can be used in immunoreaction screening for presence of
 CC anti-HGV specific antibodies, especially in patient sera, e.g. enzyme
 CC linked immunoassay (ELISA). They can also be used as vaccines, and to
 CC raise the polyclonal and monoclonal antibodies.
 XX Sequence 401 BP; 77 A; 109 C; 136 G; 79 T; 0 other;
 SQ

Query Match 84.7%; Score 14.4; DB 19; Length 401;
 Best Local Similarity 93.8%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccaccc 16
 DB 393 CATGCATTCCACCC 378

RESULT 14
 ID AAV56221/c
 XX AC AAV56221;
 XX 06-NOV-1998 (first entry)
 DT XX HGV variant 7244 DNA.
 DE XX
 XX Immunoreactive; detection; antibody; alanine aminotransferase;
 KW Non-A, Non-B, Non-C, Non-D, Non-E hepatitis virus; ds.
 XX Hepatitis G virus.
 OS US5766840-A.
 XX 16-JUN-1998.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:41:32 ; Search time 166.82 Seconds
(without alignments)
23.080 Million cell updates/sec

Title: US-09-904-420A-4

Sequence: 1 catgattaccaccacca 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCFUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	15.4	90.6	243	3	US-09-109-204-14
2	15.4	90.6	2125	3	US-09-109-204-5
C 3	14.4	84.7	289	4	US-08-991-789A-197
C 4	14.4	84.7	401	1	US-08-466-033-213
C 5	14.4	84.7	401	1	US-08-466-033-214
C 6	14.4	84.7	401	1	US-08-466-033-215
C 7	14.4	84.7	401	1	US-08-444-733-213
C 8	14.4	84.7	401	1	US-08-444-733-214
C 9	14.4	84.7	401	1	US-08-444-733-215
C 10	14.4	84.7	401	2	US-08-464-134-213
C 11	14.4	84.7	401	2	US-08-464-134-214
C 12	14.4	84.7	401	2	US-08-464-134-215
C 13	14.4	84.7	401	2	US-08-461-361-213
C 14	14.4	84.7	401	2	US-08-461-361-214
C 15	14.4	84.7	401	2	US-08-461-361-215
C 16	14.4	84.7	401	2	US-08-485-910-213
C 17	14.4	84.7	401	2	US-08-485-910-214
C 18	14.4	84.7	401	2	US-08-485-910-215
C 19	14.4	84.7	402	1	US-08-466-033-202
C 20	14.4	84.7	402	1	US-08-444-733-202
C 21	14.4	84.7	402	2	US-08-464-134-202
C 22	14.4	84.7	402	2	US-08-461-361-202
C 23	14.4	84.7	402	2	US-08-485-910-202
C 24	14.4	84.7	436	1	US-08-580-038-78
C 25	14.4	84.7	632	1	US-08-580-038-63
C 26	13.8	81.2	72	2	US-08-821-619-4
C 27	13.4	78.8	1420	1	US-08-266-414-1

C	28	13.4	78.8	1494	2	US-08-632-470-49	Sequence 49, Appl
	29	13.4	78.8	1683	3	US-08-983-045-3	Sequence 3, Appli
	30	13.4	78.8	2991	1	US-08-324-977-49	Sequence 49, Appl
	31	13.4	78.8	2991	2	US-08-384-616-49	Sequence 49, Appl
	32	13.4	78.8	2991	2	US-08-904-686A-49	Sequence 49, Appl
	33	13.4	78.8	2991	4	US-09-315-850-49	Sequence 49, Appl
	34	13.4	78.8	7863	1	US-08-324-977-35	Sequence 35, Appl
	35	13.4	78.8	7863	2	US-08-384-616-35	Sequence 35, Appl
	36	13.4	78.8	7863	2	US-08-904-686A-35	Sequence 35, Appl
	37	13.4	78.8	7863	4	US-09-315-850-35	Sequence 35, Appl
	38	13.4	78.8	7917	1	US-08-324-977-31	Sequence 31, Appl
	39	13.4	78.8	7917	2	US-08-384-616-31	Sequence 31, Appl
	40	13.4	78.8	7917	2	US-08-904-686A-31	Sequence 31, Appl
	41	13.4	78.8	7917	4	US-09-315-850-31	Sequence 31, Appl
	42	13.4	78.8	9030	1	US-08-324-977-13	Sequence 13, Appl
	43	13.4	78.8	9030	2	US-08-384-616-13	Sequence 13, Appl
	44	13.4	78.8	9030	2	US-08-904-686A-13	Sequence 13, Appl
	45	13.4	78.8	9030	4	US-09-315-850-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-109-204-14
; Sequence 14, Application US/09109204
; Patent No. 6060250
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Corley, Neil C.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN TRANSFERASES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyle Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/109,204
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0546 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT09
; CLONE: 2150892H1

US-09-109-204-14

Query Match 90.6%; Score 15.4; DB 3; Length 243;
Best Local Similarity 94.1%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccaccacca 17

Db 121 CATCCATTACCCACCCA 137

RESULT 2

US-09-109-204-5
; Sequence 5, Application US/09109204

; Patent No. 6060250

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Guegler, Karl J.

; APPLICANT: Gorgone, Gina A.

; APPLICANT: Corley, Neil C.

; APPLICANT: Patterson, Chandra

; TITLE OF INVENTION: HUMAN TRANSFERASES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/109,204

; FILING DATE: HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0546 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-855-0572

; TELEX:

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2125 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: BRAINOT09

; CLONE: 2150892

US-09-109-204-5

Query Match 90.6%; Score 15.4; DB 3; Length 2125;
Best Local Similarity 94.1%; Pred. No. 12;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccaccacca 17

Db 446 CATCCATTACCCACCCA 462

RESULT 3

US-08-991-789A-197/c

; Sequence 197, Application US/08991789A

; Patent No. 6225054

; GENERAL INFORMATION:

; APPLICANT: Fridakis, Tony N.

; Smith, John M.

; Reed, Steven G.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TREATMENT AND DIAGNOSIS OF BREAST CANCER

; NUMBER OF SEQUENCES: 292

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed IP Law Group

; STREET: 701 Fifth Avenue, Suite 6300

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/991,789A

; FILING DATE: 11-Dec-1997

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Potter, Jane E. R.

; REGISTRATION NUMBER: 33,332

; REFERENCE/DOCKET NUMBER: 210121.419C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 197:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 289 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 197:

US-08-991-789A-197

Query Match 84.7%; Score 14.4; DB 4; Length 289;
Best Local Similarity 93.8%; Pred. No. 34;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgcattaccaccacca 17

Db 77 ATACATTACCCACCCA 62

RESULT 4

US-08-466-033-213/c

; Sequence 213, Application US/08466033

; Patent No. 5766840

; GENERAL INFORMATION:

; APPLICANT: Kim, Jungsu P.

; APPLICANT: Wages, John

; APPLICANT: Young, LaVonne M.

; APPLICANT: Fry, Kirk E.

; APPLICANT: Linnen, Jeffrey M.

; TITLE OF INVENTION: Hepatitis G Virus and Molecular

; Cloning Thereof

; NUMBER OF SEQUENCES: 277

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dehlinger & Associates

; STREET: 350 Cambridge Ave., Suite 250

; CITY: Palo Alto

; STATE: CA

; COUNTRY: .USA

	NUMBER OF SEQUENCES:	277
	CORRESPONDENCE ADDRESS:	
	ADDRESSEE:	Dehlinger & Associates
	STREET:	350 Cambridge Ave., Suite 250
	CITY:	Palo Alto
	STATE:	CA
	COUNTRY:	USA
	ZIP:	94306
	COMPUTER READABLE FORM:	
	MEDIUM TYPE:	Floppy disk
	COMPUTER:	IBM PC compatible
	OPERATING SYSTEM:	PC-DOS/MS-DOS
	SOFTWARE:	PatentIn Release #1.0, Version #1.25
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER:	US/08/466,033
	FILING DATE:	
	CLASSIFICATION:	435
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/389,886
	FILING DATE:	15-FEB-1995
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/357,509
	FILING DATE:	16-DEC-1994
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/329,729
	FILING DATE:	26-OCT-1994
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/344,271
	FILING DATE:	23-NOV-1994
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/285,558
	FILING DATE:	03-AUG-1994
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	US 08/246,985
	FILING DATE:	20-MAY-1994
	ATTORNEY/AGENT INFORMATION:	
	NAME:	Fabian, Gary R.
	REGISTRATION NUMBER:	33,875
	REFERENCE/DOCKET NUMBER:	4600-0201.36/G100P11
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE:	(415) 324-0880
	TELEFAX:	(415) 324-0960
	INFORMATION FOR SEQ ID NO:	213:
	SEQUENCE CHARACTERISTICS:	
	LENGTH:	401 base pairs
	TYPE:	nucleic acid
	STRANDEDNESS:	double
	TOPOLOGY:	unknown
	MOLECULE TYPE:	DNA
	HYPOTHETICAL:	NO
	ANTI-SENSE:	NO
	ORIGINAL SOURCE:	
	INDIVIDUAL ISOLATE:	HGV-7244 VARIANT
	US-08-466-033-213	
	Query Match	84.7%; Score 14.4; DB 1;
	Best Local Similarity	93.8%; Pred. No. 35;
	Matches	15: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 catgcattaccaccc 16	
DB	393 CATGCATTTCCACACC 378	
RESULT	5	
US-08-466-033-214/c		
; Sequence 214, Application US/08466033		
; Patent No. 5766840		
GENERAL INFORMATION:		
APPLICANT:	Kim, Jungsuh P.	
APPLICANT:	Wages, John	
APPLICANT:	Young, Lavonne M.	
APPLICANT:	Fry, Kirk E.	
APPLICANT:	Linnen, Jeffrey M.	
TITLE OF INVENTION:	Hepatitis G Virus and Molecular Cloning Thereof	
TITLE OF INVENTION:	Cloning Thereof	
QY	1 catgcattaccaccc 16	
DB	393 CATGCATTTCCACACC 378	
Query Match	84.7%; Score 14.4; DB 1;	
Best Local Similarity	93.8%; Pred. No. 35;	
Matches	15: Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 catgcattaccaccc 16	
DB	393 CATGCATTTCCACACC 378	
Query Match	84.7%; Score 14.4; DB 1;	
Best Local Similarity	93.8%; Pred. No. 35;	
Matches	15: Conservative 0; Mismatches 0; Mismatches 1;	
QY	1 catgcattaccaccc 16	
DB	393 CATGCATTTCCACACC 378	
RESULT	6	
US-08-466-033-215/c		
; Sequence 215, Application US/08466033		
; Patent No. 5766840		
GENERAL INFORMATION:		
APPLICANT:	Wages, John	
APPLICANT:	Young, Lavonne M.	
APPLICANT:	Fry, Kirk E.	
APPLICANT:	Linnen, Jeffrey M.	
TITLE OF INVENTION:	Hepatitis G Virus and Molecular Cloning Thereof	
TITLE OF INVENTION:	Cloning Thereof	

APPLICANT: Kim, Jungsuh P.
APPLICANT: Wages, John
APPLICANT: Young, LaVonne M.
APPLICANT: Fry, Kirk E.
APPLICANT: Linnen, Jeffrey M.
TITLE OF INVENTION: Hepatitis G Virus and Molecular
TITLE OF INVENTION: Cloning Thereof
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,033
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 215:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: HGV-K30 VARIANT
US-08-466-033-215

Query Match 84.7%; Score 14.4; DB 1; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Caps 0;
Qy 1 catgcattaccaccc 16
Db 393 CATGCATTTCCACCC 378

RESULT 7
US-08-444-733-213/c
; Sequence 213. Application US/08444733
; Patent No. 5824507
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/444,733
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-7244 VARIANT
; US-08-444-733-213

Thu Feb 21 07:25:54 2002

us-09-904-420a-4.rni

```
Query Match      84.7%; Score 14.4; DB 1; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 catgcattaccacc 16
    ||||| |||||
Db 393 CATGCATTCCCAACC 378

RESULT 8
US-08-444-733-214/c
: Sequence 214, Application US/08444733
: Patent No. 5824507
: GENERAL INFORMATION:
: APPLICANT: Kim, Jungsuh P.
: APPLICANT: Wages, John
: APPLICANT: Young, LaVonne M.
: APPLICANT: Fry, Kirk E.
: APPLICANT: Linnen, Jeffrey M.
: TITLE OF INVENTION: Hepatitis G Virus and Molecular
: TITLE OF INVENTION: Cloning Thereof
: NUMBER OF SEQUENCES: 277
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dehlinger & Associates
: STREET: 350 Cambridge Ave., Suite 250
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/444,733
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/389,886
: FILING DATE: 15-FEB-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/357,509
: FILING DATE: 16-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,729
: FILING DATE: 26-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/344,271
: FILING DATE: 23-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/285,558
: FILING DATE: 03-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/285,543
: FILING DATE: 03-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/246,985
: FILING DATE: 20-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Fabian, Gary R.
: REGISTRATION NUMBER: 33,875
: REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
: TELEPHONE: (415) 324-0880
: TELEFAX: (415) 324-0960
: INFORMATION FOR SEQ ID NO: 214:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 401 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
```

```
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: HGV-K27 VARIANT
US-08-444-733-214

Query Match      84.7%; Score 14.4; DB 1; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 catgcattaccacc 16
    ||||| |||||
Db 393 CATGCATTCCCAACC 378

RESULT 9
US-08-444-733-215/c
: Sequence 215, Application US/08444733
: Patent No. 5824507
: GENERAL INFORMATION:
: APPLICANT: Kim, Jungsuh P.
: APPLICANT: Wages, John
: APPLICANT: Young, LaVonne M.
: APPLICANT: Fry, Kirk E.
: APPLICANT: Linnen, Jeffrey M.
: TITLE OF INVENTION: Hepatitis G Virus and Molecular
: TITLE OF INVENTION: Cloning Thereof
: NUMBER OF SEQUENCES: 277
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dehlinger & Associates
: STREET: 350 Cambridge Ave., Suite 250
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/444,733
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/389,886
: FILING DATE: 15-FEB-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/357,509
: FILING DATE: 16-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,729
: FILING DATE: 26-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/344,271
: FILING DATE: 23-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/285,558
: FILING DATE: 03-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/285,543
: FILING DATE: 03-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/246,985
: FILING DATE: 20-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Fabian, Gary R.
: REGISTRATION NUMBER: 33,875
: REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
: TELEPHONE: (415) 324-0880
: TELEFAX: (415) 324-0960
```

```

; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-K30 VARIANT
US-08-444-733-215

Query Match      84.7%; Score 14.4; DB 1; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccc 16
   ||||| |||||
Db 393 CATGCATTCCACCC 378

RESULT 10
US-08-464-134-213/c
; Sequence 213, Application US/08464134
; Patent No. 5849532
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungshuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,134
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-7244 VARIANT
US-08-464-134-213

Query Match      84.7%; Score 14.4; DB 2; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccc 16
   ||||| |||||
Db 393 CATGCATTCCACCC 378

RESULT 11
US-08-464-134-214/c
; Sequence 214, Application US/08464134
; Patent No. 5849532
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungshuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,134
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558

```



```
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-K27 VARIANT
; US-08-464-134-214

Query Match 84.7% Score 14.4; DB 2; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccacc 16
   ||||| |||||
Db 393 CATGCATTCCACCC 378

RESULT 12
US-08-464-134-215/c
; Sequence 215, Application US/08464134
; Patent No. 5849532
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsoh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linmen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,134
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-K30 VARIANT
; US-08-464-134-215

Query Match 84.7% Score 14.4; DB 2; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 catgcattaccacc 16
   ||||| |||||
Db 393 CATGCATTCCACCC 378

RESULT 13
US-08-461-361-213/c
; Sequence 213, Application US/08461361
; Patent No. 5856134
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsoh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linmen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,361
; FILING DATE:
; CLASSIFICATION: 435
```

```

; PRIOR APPLICATION DATA: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-7244 VARIANT
; US-08-461-361-213

```

Query Match 84.7%; Score 14.4; DB 2; Length 401;
 Best Local Similarity 93.8%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 catgcattaccacccc 16
    ||||| |||||
DB 393 CATGCATTCCACCC 378

```

```

RESULT 14
US-08-461-361-214/c
; Sequence 214, Application US/08461361
; Patent No. 5856134
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,361
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,543
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HGV-K27 VARIANT
; US-08-461-361-214

```

Query Match 84.7%; Score 14.4; DB 2; Length 401;
 Best Local Similarity 93.8%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 catgcattaccacccc 16
    ||||| |||||
DB 393 CATGCATTCCACCC 378

```

```

RESULT 15
US-08-461-361-215/c
; Sequence 215, Application US/08461361
; Patent No. 5856134
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, LaVonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linnen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; TITLE OF INVENTION: Cloning Thereof
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates

```

Thu Feb 21 07:25:54 2002

STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,361
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 215:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: HGV-K30 VARIANT
US-08-461-361-215

Query Match 84.7%; Score 14.4; DB 2; Length 401;
Best Local Similarity 93.8%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 catgcattaccacc 16
Db 393 CATGCATTCCCAACC 378

Search completed: February 20, 2002, 07:44:41
Job time: 189 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 07:41:32 ; Search time 3814.84 Seconds
(without alignments)
47.886 Million cell updates/sec

Title: us-09-904-420a-4
Perfect score: 17
Sequence: 1 catgattaccaccca 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST: *
- 1: em_estfun: *
 - 2: em_esthum: *
 - 3: em_estlin: *
 - 4: em_estom: *
 - 5: em_estpl: *
 - 6: em_estba: *
 - 7: em_estro: *
 - 8: em_estov: *
 - 9: em_htc: *
 - 10: gb_est1: *
 - 11: gb_est2: *
 - 12: gb_htc: *
 - 13: gb_gss: *
 - 14: em_gss_fun: *
 - 15: em_gss_hum: *
 - 16: em_gss_inv: *
 - 17: em_gss_pln: *
 - 18: em_gss_pro: *
 - 19: em_gss_rtd: *
 - 20: em_gss_vrt: *
 - 21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	94.1	201	11	BF799331
2	16	94.1	301	11	BH262580
3	16	94.1	353	11	BG386035
4	16	94.1	762	10	AW448783
5	15.4	90.6	245	11	U81209
6	15.4	90.6	248	10	AA333897
7	15.4	90.6	251	10	AV288928
8	15.4	90.6	267	10	BE148412
9	15.4	90.6	376	11	H12165
10	15.4	90.6	396	13	AQ191755
11	15.4	90.6	443	10	AA434578
12	15.4	90.6	456	13	AQ211495

13	15.4	90.6	457	10	BE543192
14	15.4	90.6	458	13	AQ697320
15	15.4	90.6	465	13	AQ774999
16	15.4	90.6	497	10	AA401238
17	15.4	90.6	515	10	A7790256
18	15.4	90.6	548	13	AQ335055
19	15.4	90.6	583	13	BH077489
20	15.4	90.6	607	13	AQ997959
21	15.4	90.6	627	10	A1799069
22	15.4	90.6	629	10	AW965401
23	15.4	90.6	634	10	BE402181
24	15.4	90.6	643	10	A1798980
25	15.4	90.6	674	11	B1308087
26	15.4	90.6	683	10	BE414229
27	15.4	90.6	713	13	AQ482035
28	15.4	90.6	781	10	AW166135
29	15.4	90.6	934	13	AZ540510
30	15.4	90.6	1031	13	CNS0010M
31	15	88.2	258	10	A1420628
32	15	88.2	286	10	BE162584
33	15	88.2	308	10	A1219879
34	15	88.2	387	13	AO507918
35	15	88.2	410	11	H13489
36	15	88.2	418	11	BG066027
37	15	88.2	447	10	BE707271
38	15	88.2	455	10	A1765616
39	15	88.2	477	10	AW881013
40	15	88.2	490	10	BE844683
41	15	88.2	509	11	BF056749
42	15	88.2	521	10	AW473180
43	15	88.2	521	13	AZ082044
44	15	88.2	530	13	AZ123706
45	15	88.2	534	10	A1633707

ALIGNMENTS

RESULT 1

BF799331 201 bp mRNA EST 12-JAN-2001
LOCUS RC6-C10122-091000-021-D10 C10122 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF799331
VERSION BF799331.1 GI:12128320
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 201)

AUTHORS

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,

Brustein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shoiguon sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC6&t2=RC6-C10122-091000-021-D10&t3=2000-10-09&t4=1>)

Seq primer: puc 18 forward

Thu Feb 21 07:25:57 2002

High quality sequence start: 16
High quality sequence stop: 201.
Location/Qualifiers
1..201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0122"
/dev_stage="Adult"
/note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
45 a 71 c 50 g 35 t

BASE COUNT
ORIGIN

Query Match 94.1%; Score 16; DB 11; Length 201;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 catcattaccaccc 16
|||||
Db 138 CATGATTACCCACCC 153

RESULT 2
BB262580 301 bp mRNA EST 07-JUL-2000
LOCUS BB262580 RIKEN full-length enriched, 10 days neonate cortex Mus
DEFINITION musculus cDNA clone A830007A11 3', mRNA sequence.
ACCESSION BB262580
VERSION BB262580.1 GI:8959035
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 301)
Komo.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci
P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,
Hirozane.T., Hori.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,
Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,
Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C., Kusakabe.M.,
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K., Shibata.K., Shibata
Y., Shigemoto.Y., Shinagawa.A., Shiraki.T., Sogabe.Y., Sugahara.Y.,
Suzuki.H., Suzuki.H., Tagawa.A., Takahashi.F., Tominaga.N., Toya
T., Tsunoda.Y., Watahiki.A., Yokota.T., Yoshida.K., Yoshiki.A., Yoshino
M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC) Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci.P., Nishiyama.Y., Westover.A., Itoh.M., Nagaoaka.S., Sasaki
N., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh.M., Katsunai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,
Tomaru.Y., Carninci.P., Shibata.Y., Ozawa.Y., Muramatsu.M., Okazaki
Y. and Hayashizaki.Y.
Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)
Carninci.P. and Hayashizaki.Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..301
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="A830007A11"
/clone_lib="RIKEN full-length enriched, 10 days neonate
cortex"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SmaI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATATCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I.
69 a 115 c 47 g 70 t

BASE COUNT
ORIGIN

Query Match 94.1%; Score 16; DB 10; Length 301;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 atgattaccaccc 17
|||||
Db 30 ATGCATTACCCACCA 45

RESULT 3
BG386035 353 bp mRNA EST 12-MAR-2001
LOCUS BG386035
DEFINITION BG386035 Homo sapiens cDNA clone IMAGE:4583473 5',
mRNA sequence.
ACCESSION BG386035
VERSION BG386035.1 GI:13279481
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 353)
NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI307 row: h column: 02
High quality sequence stop: 276.

```

FEATURES
  source      Location/Qualifiers
1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EORI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      80 a 111 c 122 g 40 t
ORIGIN

Query Match      94.1%; Score 16; DB 11; Length 353;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgcattaccaccc 16
|||||
Db 240 CATGCATTACCCACCC 255

RESULT 4
AM448783/c
LOCUS      AM448783      762 bp      mRNA      EST      03-JAN-2001
DEFINITION BRY_1422 BRY Triticum aestivum cDNA clone P35-1P, mRNA sequence.
ACCESSION  AM448783
VERSION     AM448783.1 GI:12019318
KEYWORDS   EST.
SOURCE     Bread wheat.
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
            1 (bases 1 to 762)
REFERENCE   Clarke,B.C., Hobbs,M. and Appels,R.
AUTHORS    Genes active in developing wheat endosperm
            Unpublished (2000)
JOURNAL
COMMENT    Contact: Bryan Clarke
            Division of Plant Industry
            C.S.I.R.O.
            GPO Box 1600, Canberra, ACT, Australia
            Tel: 61 2 6246 5054
            Fax: 61 2 6246 5000
            Email: bryan@dpi.csiro.au.
            Location/Qualifiers
FEATURES
  source      1. .762
/organism="Triticum aestivum"
/cultivar="Wyuna"
/db_xref="taxon:4565"
/clone="P35-1P"
/cell_type="endosperm"
BASE COUNT      205 a 196 c 171 g 179 t 11 others
ORIGIN

Query Match      94.1%; Score 16; DB 10; Length 762;
Best Local Similarity 94.1%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccca 17
|||||
Db 579 CATGCATTACCCACNCA 563

FEATURES
  source      Location/Qualifiers
1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EORI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      80 a 111 c 122 g 40 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 11; Length 245;
Best Local Similarity 94.1%; Pred. No. 1.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccca 17
|||||
Db 73 CATGCATTACCCACCCA 89

RESULT 6
AA333897
LOCUS      AA333897      248 bp      mRNA      EST      21-APR-1997
DEFINITION EST38225 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION  AA333897
VERSION     AA333897.1 GI:1986317
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 248)
REFERENCE   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
AUTHORS    ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Colton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodok,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moren-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Jii,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
            Hu,J.S., Greene,J.M., Gruher,J.J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., HungJun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,

```


AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

TITLE

Sequence tags

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=MR0-HT0241-131

299-002-g08&t3=1999-12-13&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 33

High quality sequence stop: 267.

Location/Qualifiers

1. .267

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0241"

/dev_stage="Adult"

/note="Organ: head,neck; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

72 a 54 c 62 g 79 t

BASE COUNT

ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 267;

Best Local Similarity 94.1%; Pred. No. 1.7e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17

|||||

Db 244 CATGCATTACCAACCA 260

|||||

RESULT 9

H12165

LOCUS

DEFINITION

IMAGE:47965 5', mRNA sequence.

H12165

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 376)

Miller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,

Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,F., Waterston

,R., Williamson,A., Wohlmann,P. and Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1841

High quality sequence stops: 339

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 1841 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 339.

Location/Qualifiers

1. .376

/organism="Homo sapiens"

/db_xref="GDB:420506"

/db_xref="taxon:9606"

/clone_lib="IMAGE:47965"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: Lfamid BA; Site_1: Not

I; Site_2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer [5,

AACTGAGAGATTCCGCCGACAGGAATTTTTCCTTTT 3';

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the Lfamid BA vector.

Library went through one round of normalization. Library

constructed by Benito Soares and M.Fatima Bonaldo."

118 a 87 c 68 g 101 t

BASE COUNT

ORIGIN

Query Match 90.6%; Score 15.4; DB 11; Length 376;

Best Local Similarity 94.1%; Pred. No. 1.7e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17

|||||

Db 63 CATGCATTACCAACCA 79

|||||

RESULT 10

AQ191755/c

LOCUS

DEFINITION

sapiens genomic clone Plate=3228 Col=20 Row=K, DNA sequence.

AQ191755

ACCESSION

VERSION

KEYWORDS

SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 396)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3228 row: K column: 20

Class: BAC ends

High quality sequence stop: 396.
 Location/Qualifiers
 1. .396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="plate=3228 Col=20 Row=K"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

137 a 81 c 81 g 94 t 3 others

BASE COUNT
 ORIGIN

Query Match 90.6%; Score 15.4; DB 13; Length 396;
 Best Local Similarity 94.1%; Pred. No. 1.7e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17
 ||| ||||| ||||| |||||

Db 333 CATGCATTACCCCTCCA 317

RESULT 11
 AA434578/c
 LOCUS
 DEFINITION
 zw52b08.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
 IMAGE:773655 5' similar to TR:G1050752 G1050752
 KYNURENINE/ALPHA-AMINOADIPATE AMINOTRANSFERASE ;, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 443)
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
 Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,
 T., Waterston,R. and Wilson,R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estwatson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28mi3 rev2 ET from Amersham
 High quality sequence stop: 372.
 Location/Qualifiers
 1. 443
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:773655"
 /clone_lib="Soares_total_fetus_Nb2HF8_9w"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /note="Vector: pT7D3-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not J; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from pooled 8-9 week
 (total) fetus material with a Not I - oligo(dT) primer [5'
 TGTTACCAATCGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

121 a 89 c 98 g 135 l

BASE COUNT

ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 443;
 Best Local Similarity 94.1%; Pred. No. 1.7e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17
 ||| ||||| ||||| |||||

Db 252 CATGCATTACCCACCCA 236

RESULT 12
 AQ211495
 LOCUS
 DEFINITION
 HS_3243_A1_B05_MR CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3243 Col=9 Row=C, DNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 456)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3243 row: C column: 9
 Class: BAC ends
 High quality sequence stop: 456.
 Location/Qualifiers
 1. 456
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate=3243 Col=9 Row=C"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

128 a 96 c 74 g 150 t 8 others

BASE COUNT
 ORIGIN

Query Match 90.6%; Score 15.4; DB 13; Length 456;
 Best Local Similarity 94.1%; Pred. No. 1.7e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccacca 17
 ||||| ||||| ||||| |||||

Db 261 CATGCATTACCCACCCA 277

RESULT 13
 BF543192
 LOCUS
 DEFINITION
 BF543192 457 bp mRNA EST 09-AUG-2000
 601069175F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:345251 5',
 mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 EST.

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 457)
AUTHORS NIH-MGC <http://mgc.ncbi.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENTS

DNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LUNL at:
<http://image.llnl.gov>

```

http://anage.irs.ri.gov:
Plate: LIA88441 row: f column: 20
High quality sequence start: 7
High quality sequence stop: 429.
Location/Qualifiers
FEATURES

```

```
1. .457
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3455251"
```

```
/clone_lib="-NIH_MGC_c12"  
/tissue_type="cervical carcinoma cell line"  
/lab_host="DH10B"  
/note="Organ; Vector: pCMV-SPORT6; Site_1: NotI;  
sites: SalI; cloned unidirectionally primer: Oligo dt.
```

BASE COUNT
ORIGIN

Query Match	90.6%	Score 15.4;	DB 10;	Length 457;
Best Local Similarity	94.1%	Pred. No. 1.7e+03;		

[illegible]

DB 112 CATCCATTACCCACGCCA 128

RESULT 14

NC607320

	A0697320	458 bp	DNA	GSS
LOCUS	HS_5531_B1_E07_T7A_RPCI-11		Human Male BAC Library	06-JUL-1999
DEFINITION	genomic clone Plate=1107 Col=13 Row=J,		DNA sequence.	
ACCESSION	A0697320			

VERSION	AQ697320.1	GI:5387568
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
 Hominidae; Homo.
 1 (bases 1 to 458)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 and
 AUTHORS
 REFERENCE

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 Hood, L.
 Kellier, A., Shaker, R., Furlong, J., Young, R., Zink, S., Andrews, B., and
 proc Natl Acad Sci U S A. 96 (17), 9739-9744 (1999)
JOURNAL

99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPC1-11. For BAC

Thu Feb 21 07:25:57 2002

```

/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 126 a 87 c 81 g 166 t 5 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 13; Length 465;
Best Local Similarity 94.1%; Pred. No. 1.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgcattaccaccaca 17
   ||||| |||||
Db 362 CATGCATTACCACCCA 378
```

Search completed: February 20, 2002, 10:25:41
Job time: 9849 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:25:50 ; Search time 1729.86 Seconds
(without alignments)
333.785 Million cell updates/sec

Title: US-09-904-420a-1_COPY_882_916
Perfect score: 35
Sequence: 1 cccctggagcaggcagcagcccccagccagc 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_om.*
20: em_of.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vi.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	35	100.0	1115	9	HSTCFID	Z47361 H.sapiens T
2	35	100.0	1165	9	HSTCFIE	Z47362 H.sapiens T
3	35	100.0	1254	9	HSTCFIA	X59869 Human TCF-1
4	35	100.0	2814	9	HSTCFIB	X59870 Human TCF-1
5	35	100.0	2855	9	HSTCFIG	X63901 Homo sapien
6	35	100.0	2910	9	HSTCFIC	X59871 Human TCF-1
c 7	35	100.0	84544	2	AC009012	AC009012 Homo sapi
c 8	35	100.0	86914	2	AC011336	AC011336 Homo sapi
c 9	35	100.0	159420	2	AC009017	AC009017 Homo sapi
10	35	100.0	200831	2	AC008608	AC008608 Homo sapi
11	26.6	76.0	1790	10	MMTCSTF	X61385 Mouse mRNA
12	25	71.4	162496	2	AC044846	AC044846 Mus muscu
13	24.4	69.7	126712	9	AL137077	AL137077 Human DNA
14	23.2	66.3	1500	9	AF034801	AF034801 Homo sapi
c 15	23.2	66.3	200794	2	AL451082	AL451082 Homo sapi
c 16	22.8	65.1	3285	5	U00111	U00111 Gallus gall
17	22.8	65.1	77022	9	HS983L19	AL023802 Human DNA
c 18	22.8	65.1	81210	2	AC019242	AC019242 Homo sapi
c 19	22.8	65.1	136747	2	AC074389	AC074389 Homo sapi
c 20	22.8	65.1	163338	2	AC011603	AC011603 Homo sapi
c 21	22.8	65.1	175793	9	AC005920	AC005920 Homo sapi
22	22.8	65.1	180110	2	AC073610	AC073610 Homo sapi
c 23	22.8	65.1	180749	2	AC025256	AC025256 Homo sapi
24	22.6	64.6	3935	9	AY013874	AY013874 Homo sapi
25	22.6	64.6	4506	9	AB028971	AB028971 Homo sapi
26	22.6	64.6	4554	9	AB037821	AB037821 Homo sapi
27	22.6	64.6	54852	2	AC090367	AC090367 Homo sapi
28	22.6	64.6	87735	2	AC072045	AC072045 Homo sapi
29	22.6	64.6	146450	2	AC022497	AC022497 Homo sapi
30	22.6	64.6	162836	2	AC021594	AC021594 Homo sapi
31	22.6	64.6	170803	2	AP001641	AP001641 Homo sapi
32	22.6	64.6	172120	2	AC022337	AC022337 Homo sapi
33	22.6	64.6	190030	2	AC079121	AC079121 Homo sapi
c 34	22.2	63.4	1812	9	AB040672	AB040672 Homo sapi
c 35	22.2	63.4	2535	9	AB050509	AB050509 Macaca fa
36	22.2	63.4	3560	1	AF218066	AF218066 Streptomy
c 37	22.2	63.4	155742	9	CNS01DMG	AF137190 Human chr
c 38	22.2	63.4	157546	2	AC048368	AC048368 Homo sapi
c 39	22.2	62.9	40374	3	LMFL3302	AL359781 Lelshmani
40	21.8	62.3	351	6	AX073564	AX073564 Sequence
41	21.8	62.3	567	6	AX059849	AX059849 Sequence
42	21.8	62.3	823	6	AX059857	AX059857 Sequence
43	21.8	62.3	1096	9	AF338109	AF338109 Homo sapi
44	21.8	62.3	1894	4	AB011673	AB011673 Ros tauru
45	21.8	62.3	1986	6	AX073574	AX073574 Sequence

ALIGNMENTS

RESULT 1

HSTCFID	HSTCFID	1115 bp	mRNA	PRI	09-JAN-1995
LOCUS	H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.				
DEFINITION	Z47361				
ACCESSION	247361.1	GI:61988J			
VERSION	splice form D; T cell factor 1.				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	1 (bases 1 to 1115)				
AUTHORS	Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.				
TITLE	The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon 1A				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1115)				
AUTHORS	van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D., van der Kleij, R., Geurts van Kessel, A. and Clevers, H.				

TITLE The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
REFERENCE 92235082
MEDLINE 3 (bases 1 to 1115)
AUTHORS Ballhausen, W.G.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES
source Location/Qualifiers
1. .1115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1115
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1115
/gene="TCF-1"
2. .754
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form D"
/protein_id="CAA87439.1"
/db_xref="GI:619882"
/translation="MYKETVYSAFNLMHYPPSGAGQHQPQPLHKANOPPHGVPO
LSLIEFNSPHPTAPADISQVHRPLQTPDLSGFVSLTSGSMGQLPHTVSMFTHPS
LMLGSGVGPHPAAIPHPAIVPPSGQELQPDRLNKTQAEKAEKAKKPTIKKPLNA
FMLYKEMRAKVIATECTLKESAAINQILGRWIALSREEQAKYTELARKERQILHMOLY
PWMSARDNYGKKRRSREKHQESTTDSNLSHS"
731. .751
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1D"
BASE COUNT 272 a 407 c 265 g 171 t
ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 1115;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccctcgggagcggcgagcaccgccgagcgagc 35
|||||
Db 55 CCCCTCGGAGCAGCGCAGCACCAGCGCGCAGC 89

RESULT 2
HSTCF1E
LOCUS HSTCF1E 1165 bp mRNA PRI 09-JAN-1995
DEFINITION H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.
ACCESSION Z47362
VERSION 247362.1 GI:619883
KEYWORDS splice form E; T cell factor 1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1165)
AUTHORS Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.
TITLE The human high mobility group (HMG)-box transcription factor TCF-1:
novel isoforms due to alternative splicing and usage of a new exon
IXA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1165)
AUTHORS van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D.,

TITLE The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
REFERENCE 92235082
MEDLINE 3 (bases 1 to 1165)
AUTHORS Ballhausen, W.G.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES
source Location/Qualifiers
1. .1165
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1. .1165
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1. .1165
/gene="TCF-1"
2. .1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form F"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKETVYSAFNLMHYPPSGAGHQPQPLHKANOPPHGVPO
LSLIEFNSPHPTAPADISQVHRPLQTPDLSGFVSLTSGSMGQLPHTVSMFTHPS
LMLGSGVGPHPAAIPHPAIVPPSGQELQPDRLNKTQAEKAEKAKKPTIKKPLNA
FMLYKEMRAKVIATECTLKESAAINQILGRWIALSREEQAKYTELARKERQILHMOLY
PWMSARDNYGKKRRSREKHQESTTDPGSPKKCARFGLNGQOTDMCGPCRKKKCIY
LPGEGRCPSPVPSDPSALCGSPAPQSPSYHLPRPTELLTSPAEIRHLHPQVSPL
LSASQPGPHRPAPAPCAHRYSNRNLDRWPSKRTTGRLEQETP"
803. .1141
mat_peptide
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccctcgggagcggcgagcaccgccgagcgagc 35
|||||
Db 55 CCCCTCGGAGCAGCGCAGCACCAGCGCGCAGC 89

RESULT 3
HSTCF1A
LOCUS HSTCF1A 1254 bp mRNA PRI 14-JUN-1991
DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form A).
ACCESSION X59869
VERSION X59869.1 GI:36785
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1254)
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

REFERENCE AUTHORS TITLE	The Netherlands		JOURNAL MEDLINE COMMENT FEATURES	EMBO J. 10 (1), 123-132 (1991)		
	2 (bases 1 to 1254)			91114695		
	van de Wetering, M., Oosterwegel, M., Dooljes, D. and Clevers, H.			See also X59869-X59871.		
	Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box			Location/Qualifiers		
JOURNAL MEDLINE COMMENT FEATURES	source		misc_feature	1..2814		
	EMBO J. 10 (1), 123-132 (1991)			/organism="Homo sapiens"		
	91114695			/db_xref="taxon:9606"		
	See also X59869-X59871.			/cell_type="T lymphocyte"		
source	Location/Qualifiers		misc_feature	/cell_line="Jurkat and HPB-ALL"		
	1..1254			/clone_lib="cDNA"		
	/organism="Homo sapiens"			1..2814		
	/db_xref="taxon:9606"			/evidence="experimental"		
misc_feature	80..889		misc_feature	1..2814		
	/product="T cell factor 1, splice form A"			/gene="TCF-1"		
	/protein_id="CAA42526.1"			80..889		
	/db_xref="GI:36786"			/gene="TCF-1"		
misc_feature	539..769		misc_feature	/codon_start=1		
	/translation="MYKETVSAFNLMLHYPSPGAGQHPOQPPLHKANQPPHGVPO LSLYHFNSPHPTAPADISQKVHRPLOTDLSCFYSLTSGSMQPLPHVSWFTHPS LMGSGVPGHPAAIPHPAIVPPSGKQELQPDRLNLTQAEKAEKAKPTIKKPLNA FMLYKMKRAXVIAECTLKESAAINQILGRWHALSREQAKYELAKKERQJLMOLY PGWSARDNYGKKRRSRKXKHQESTTETNPRELKDGNGQESLSMSSSSPA"			/product="T cell factor 1 splice form B"		
	539..769			/protein_id="CAA42527.1"		
	/db_xref="GI:36788"			/db_xref="GI:36788"		
misc_feature	810		misc_feature	/db_xref="SWISS-PROT:P36402"		
	/gene="TCF-1"			/translation="MYKETVSAFNLMLHYPSPGAGQHPOQPPLHKANQPPHGVPO LSLYHFNSPHPTAPADISQKVHRPLOTDLSCFYSLTSGSMQPLPHVSWFTHPS LMGSGVPGHPAAIPHPAIVPPSGKQELQPDRLNLTQAEKAEKAKPTIKKPLNA FMLYKMKRAXVIAECTLKESAAINQILGRWHALSREQAKYELAKKERQJLMOLY PGWSARDNYGKKRRSRKXKHQESTTETNPRELKDGNGQESLSMSSSSPA"		
	/note="alternative splice site"			/db_xref="SWISS-PROT:P36402"		
	/note="HMG box"			/translation="MYKETVSAFNLMLHYPSPGAGQHPOQPPLHKANQPPHGVPO LSLYHFNSPHPTAPADISQKVHRPLOTDLSCFYSLTSGSMQPLPHVSWFTHPS LMGSGVPGHPAAIPHPAIVPPSGKQELQPDRLNLTQAEKAEKAKPTIKKPLNA FMLYKMKRAXVIAECTLKESAAINQILGRWHALSREQAKYELAKKERQJLMOLY PGWSARDNYGKKRRSRKXKHQESTTETNPRELKDGNGQESLSMSSSSPA"		
BASE COUNT ORIGIN	320 a 391 c 335 g 208 t		BASE COUNT ORIGIN	763 a 831 c 648 g 572 t		
	Query Match 100.0%; Score 35; DB 9; Length 1254;			Query Match 100.0%; Score 35; DB 9; Length 2814;		
	Best Local Similarity 100.0%; Pred. No. 0.039;			Best Local Similarity 100.0%; Pred. No. 0.031;		
	Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1 cccctcgaggagcagccagccagccagccagc 35		Oy	1 cccctcgaggagcagccagccagccagccagc 35		
	Db 133 CCCCTCGGAGCAGGCGAGCAGCCAGCCGAGC 167			Db 133 CCCCTCGGAGCAGGCGAGCAGCCAGCCGAGC 167		
RESULT 4	HSTCF1B		RESULT 5	HSTCF1G		
	LOCUS			LOCUS		
	DEFINITION			DEFINITION		
	ACCESSION			ACCESSION		
REFERENCE	X59870.1 GI:36787		REFERENCE	X63901.1 GI:36791		
	DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.			T-cell transcription factor; transcription factor.		
	human.			human		
	ORGANISM			ORGANISM		
REFERENCE	Homo sapiens		REFERENCE	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 2814)			1 (bases 1 to 2855)		
AUTHORS	van de Wetering, M.		AUTHORS	van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D., Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.		
	Direct Submission			The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization		
	Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht, The Netherlands			J. Biol. Chem. 267 (12), 8530-8536 (1992)		
	2 (bases 1 to 2814)			2 (bases 1 to 2855)		
TITLE	Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box		TITLE	Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties		
	Mol. Cell. Biol. 16 (3), 745-752 (1996)			Mol. Cell. Biol. 16 (3), 745-752 (1996)		
	96182076			96182076		
	3 (bases 1 to 2855)			3 (bases 1 to 2855)		
JOURNAL MEDLINE AUTHORS	Van de Wetering, M.L.		JOURNAL MEDLINE AUTHORS	Van de Wetering, M.L.		

TITLE	Direct Submission	Submitted (07-JAN-1992)	M.L. Van De Wetering, Department of Immunology, P.O. Box 8550, 3508 GA Utrecht, THE NETHERLANDS
FEATURES	location/Qualifiers		
source	1. .2855		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/cell_type="T-lymphocyte"		
	/clone="phi-TCF-[5,8,11]"		
	/map="5q31.1"		
exon	829..924		
	/gene="TCF-1"		
	/number=1		
CDS	join(829..924,970..1074,1120..1209,1255..1374,1420..1581,1691..1798,1843..1890,2054..2134)		
	/gene="TCF-1"		
	/codon_start=1		
	/protein_id="CAB56795.1"		
	/db_xref="GI:6006365"		
	/translation="MYKTVYSAFNLLMHYPPSGAGQHQPQPLHKANOPPHGVPO LSLVEHNSPHPTTAPADISQKVRHPLQTDLSGFYSLTSGSMGQLPHTVSWFTPHS LMLGSGVPGHPAAIPHPAIVPPSGKQELQPDFRNLKTAQSAKAEKAKKPTTKKPLNA FMLYGEMRAKYTAETCLKESAAINQILGRSWHALSREQAAYVELAKERQLHMQLY PGWSADNYYGKKRKSREKHQESTTETNPRELKDGNGQESLSMSSSSSPA"		
gene	join(829..924,970..1074,1120..1209,1255..1374,1420..1581,1691..1798,1843..1890,2054..2134)		
	/gene="TCF-1"		
	925..970		
intron	/number=1		
exon	970..1074		
	/gene="TCF-1"		
	/number=2		
intron	1075..1119		
	/number=2		
exon	1120..1209		
	/gene="TCF-1"		
	/number=3		
intron	1210..1254		
	/number=3		
exon	1255..1374		
	/gene="TCF-1"		
	/number=4		
intron	1375..1419		
	/number=4		
exon	1420..1581		
	/gene="TCF-1"		
	/number=5		
intron	1582..1690		
	/number=5		
exon	1691..1798		
	/gene="TCF-1"		
	/number=6		
intron	1799..1842		
	/number=6		
exon	1843..1890		
	/gene="TCF-1"		
	/number=7		
intron	1891..2053		
	/number=1		
exon	2054..2134		
	/gene="TCF-1"		
	/number=8		
BASE COUNT	592 a 929 c 771 g 562 t 1 others		
ORIGIN			
Query Match	100.08; Score 35; DB 9; Length 2855;		
Best Local Similarity	100.08; Pred. No. 0.031;		
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 ccctcgggagcaggcagcacccccagccgacg 35		

Center clone name: CIT978SKB_113120

Summary Statistics

Consensus quality: 174733 bases at least Q40

Consensus quality: 186693 bases at least Q30

Consensus quality: 190547 bases at least Q20

Estimated insert size: 165000; pulse field gel estimation

Estimated insert size: 198031; sum-of-contigs estimation

Quality coverage: 7.77 in Q20 bases; pulse field gel estimation

Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

- * consists of 29 contigs. The true order of the pieces
- * is not known and their order in this sequence record is
- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

* 1 1115: contig of 1115 bp in length

* 1116 1215: gap of unknown length

* 1216 2567: contig of 1352 bp in length

* 2568 2667: gap of unknown length

* 2668 3745: contig of 1078 bp in length

* 3746 3845: gap of unknown length

* 3846 4920: contig of 1075 bp in length

* 4921 5020: gap of unknown length

* 5021 6044: contig of 1024 bp in length

* 6045 6144: gap of unknown length

* 6145 7285: contig of 1141 bp in length

* 7286 7385: gap of unknown length

* 7386 8659: contig of 1274 bp in length

* 8660 8759: gap of unknown length

* 8760 9804: contig of 1045 bp in length

* 9805 11696: contig of 1792 bp in length

* 11697 11796: gap of unknown length

* 11797 12841: contig of 1045 bp in length

* 12842 12941: gap of unknown length

* 12943 15043: contig of 2102 bp in length

* 15044 15143: gap of unknown length

* 15144 17289: contig of 2146 bp in length

* 17290 17389: gap of unknown length

* 17390 20332: contig of 2843 bp in length

* 20333 21884: contig of 1552 bp in length

* 21885 21984: gap of unknown length

* 21985 24278: contig of 2294 bp in length

* 24279 28103: contig of 3725 bp in length

* 28104 30867: gap of unknown length

* 30868 30967: gap of unknown length

* 30968 34914: contig of 3947 bp in length

* 34915 35014: gap of unknown length

* 35015 36030: contig of 1016 bp in length

* 36031 36130: gap of unknown length

* 36131 38852: contig of 2722 bp in length

* 38853 38952: gap of unknown length

* 38953 41871: contig of 2919 bp in length

* 41872 41971: gap of unknown length

* 41972 47080: contig of 5109 bp in length

* 47081 47180: gap of unknown length

* 47181 49750: contig of 2570 bp in length

* 49751 49850: gap of unknown length

* 49851 57118: contig of 7268 bp in length

* 57119 57218: gap of unknown length

* 57219 66046: contig of 8828 bp in length

* 66047 66146: gap of unknown length

* 66147 81047: gap of unknown length

* 81047 81147: gap of unknown length

* 81147 101650: contig of 20504 bp in length

* 101651 101750: gap of unknown length

* 101751 137108: contig of 35358 bp in length

* 137109 137208: gap of unknown length

* 137209 200831: contig of 63623 bp in length.

FEATURES

source

1. .200831

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/clone="CTB-113120"

/clone_lib="Caltech human BAC library B"

BASE COUNT 53266 a 46907 c 47140 g 50695 t 2823 others

ORIGIN

Query Match 100.0% Score 35; DB 2; Length 200831;

Best Local Similarity 100.0%; Pred. No. 0.01;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgggagcaggcagcagcccccagcagcagc 35

|||||

Db 184039 CCCCTCGGAGCAGGCAGCAGCCCGAGCGCAGC 184073

RESULT 11

MMTCSTF

LOCUS 1790 bp mRNA ROD 02-SEP-1991

DEFINITION Mouse mRNA for T-cell specific transcription factor.

ACCESSION X61385

VERSION X61385.1 GI:54760

KEYWORDS HMG box; T-cell specific transcription factor; transcription factor.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1790)

Oosterweel, M.A.

Direct Submission

Submitted (12-AUG-1991) M.A. Oosterweel, Dept of Clin Immunol., Univ Hospital Utrecht, P.O. Box 85500, 3508 GA Utrecht, THE NETHERLANDS

REFERENCE 2 (bases 1 to 1790)

Oosterweel, M., van de Metering, M., Dooljes, D., Klomp, L., Winoto, A., Georgopoulos, K., Meijlink, F. and Clevers, H.

Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers

J. Exp. Med. 173 (5), 1133-1142 (1991)

JOURNAL

MEDLINE 91217625

FEATURES

source

Location/Qualifiers

1. .1790

/organism="Mus musculus"

/isolate="M2a/M5"

/db_xref="taxon:10090"

/tissue_type="thymus"

/cell_type="T cell"

/clone_lib="lambda zap cDNA"

1. .1790

190. .1101

/codon_start=1

/product="T-cell specific transcription factor"

/protein_id="CAA43658.1"

/db_xref="GI:54761"

/db_xref="MGD:MGI:98507"

/db_xref="SWISS-PROT:Q00417"

/translation="MYKETVYSANFLMPYPPASGAGOHPOPPPLHNKPGQPPHGV

OLSPVYEHFSSPHPTAPADISOKOGVHPLOTDLSCGFYSITSGMGLPHTYVSPS

PPLPYLSPSCGGRQHPFAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP

KQELQPYDRNLKTQAEPAKAEKAPVTKKPLNAPLMLYMKEMRAKVIAECTLKSAAL

NOTIGRRWALSREQAQYIELARKERQI.HMQLYPGWSARUNYCKKKRRSRKHEQEST

TGGKRNAFGCTYPEKAAAPAPFLPMTVL"

misc_feature 489. .581

/note="ACT splice/exon M5"

misc_feature 761. .984

```

BASE COUNT      423 a      587 c      436 g      343 t      1 others
ORIGIN

Query Match      76.0%; Score 26.6; DB 10; Length 1790;
Best Local Similarity 87.9%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 cctcggagcagggcagcagcccccagcagcagc 35
        |||||
Db      245 CCTCGGAGCAGCGCAGCATCGCAGCTCAAC 277

RESULT 12
AC044846/c
LOCUS      AC044846      162496 bp      DNA      HTG      22-JUL-2000
DEFINITION Mus musculus chromosome 11 clone RP23-46J7 map 11, WORKING DRAFT
AC044846
ACCESSION      AC044846.2      GI:9369533
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 162496)
Mus musculus chromosome 11, clone RP23-46J7
Unpublished
2 (bases 1 to 162496)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balgwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodghe,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McEltrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triquillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 22, 2000 this sequence version replaced gi:7543812.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9174
Center clone name: 46_J7
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149639 bases at least Q40
Consensus quality: 156386 bases at least Q30

```

```

Consensus quality: 158956 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 160296; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N. But the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1373: contig of 1373 bp in length
* 1374 1473: gap of 100 bp
* 1474 2786: contig of 1313 bp in length
* 2787 2886: gap of 100 bp
* 2887 4262: contig of 1376 bp in length
* 4263 4362: gap of 100 bp
* 4363 6417: contig of 2055 bp in length
* 6418 6517: gap of 100 bp
* 6518 9412: contig of 2895 bp in length
* 9413 9512: gap of 100 bp
* 9513 10790: contig of 1278 bp in length
* 10791 10890: gap of 100 bp
* 10891 15080: contig of 4190 bp in length
* 15081 15180: gap of 100 bp
* 15181 18824: contig of 3644 bp in length
* 18825 18924: gap of 100 bp
* 18925 23129: contig of 4205 bp in length
* 23130 23229: gap of 100 bp
* 23230 26859: contig of 3630 bp in length
* 26860 26959: gap of 100 bp
* 26960 32575: contig of 5616 bp in length
* 32576 32675: gap of 100 bp
* 32676 38781: contig of 6106 bp in length
* 38782 38881: gap of 100 bp
* 38882 44764: contig of 5883 bp in length
* 44765 44864: gap of 100 bp
* 44865 52115: contig of 7251 bp in length
* 52116 52215: gap of 100 bp
* 52216 59927: contig of 7712 bp in length
* 59928 60027: gap of 100 bp
* 60028 67115: contig of 7088 bp in length
* 67116 67215: gap of 100 bp
* 67216 75188: contig of 7973 bp in length
* 75189 75288: gap of 100 bp
* 75289 85641: contig of 10353 bp in length
* 85642 85741: gap of 100 bp
* 85742 97849: contig of 12108 bp in length
* 97850 97949: gap of 100 bp
* 97950 109572: contig of 11623 bp in length
* 109573 109672: gap of 100 bp
* 109673 122441: contig of 12769 bp in length
* 122442 122541: gap of 100 bp
* 122542 140243: contig of 17702 bp in length
* 140244 140343: gap of 100 bp
* 140344 162496: contig of 22153 bp in length.
Location/Qualifiers
1. .162496
/organism="Mus musculus"
/db_xref="Laxon:10090"
/chromosome="11"
/map="11"
/clone_lib="RP23-46J7"
/clone="RP23-46J7"
/note="assembly_fragment"
1474. .2786
/note="assembly_fragment"
2887. 4262
/note="assembly_fragment"

```

```

FEATURES
source
misc_feature
misc_feature
misc_feature

```

misc_feature	4363..6417	/note="assembly_fragment"
misc_feature	6518..9412	/note="assembly_fragment"
misc_feature	9513..10790	/note="assembly_fragment"
	clone_end:SP6 vector_side:right"	
misc_feature	10891..15080	/note="assembly_fragment"
misc_feature	15181..18824	/note="assembly_fragment"
misc_feature	18925..23129	/note="assembly_fragment"
misc_feature	23230..26859	/note="assembly_fragment"
misc_feature	26960..32575	/note="assembly_fragment"
misc_feature	32676..38781	/note="assembly_fragment"
misc_feature	38882..44764	/note="assembly_fragment"
misc_feature	44865..52115	/note="assembly_fragment"
misc_feature	52216..59927	/note="assembly_fragment"
misc_feature	60028..67115	/note="assembly_fragment"
misc_feature	67216..75188	/note="assembly_fragment"
misc_feature	75289..85641	/note="assembly_fragment"
misc_feature	85742..97849	/note="assembly_fragment"
misc_feature	97950..109572	/note="assembly_fragment"
misc_feature	109673..122441	/note="assembly_fragment"
misc_feature	122542..140243	/note="assembly_fragment"
	clone_end:T7 vector_side:left"	
misc_feature	140344..162496	/note="assembly_fragment"
BASE COUNT	42470 a 37212 c 36404 g 44204 t	2206 others
ORIGIN		
Query Match	71.4%;	Score 25; DB 2; Length 162496;
Best Local Similarity	84.8%;	Pred. No. 26;
Matches	28; Conservative	0; Mismatches
	5; Indels	0; Gaps
Seq	3 cctcgggagcaggccagcacccccacgcgcgaac 35	
db	61980 CCTCCGAGCAGGGCGGCATCTCCGACGCCCTCAAC 61948	
RESULT	13	
LOCUS	Al137077	PRI 23-APR-2001
DEFINITION	Human DNA sequence from clone RP11-11M20 on chromosome 20 Contains the 5' part of the TAF2C1 gene for TATA box binding protein (TBP)-associated factor C1 (RNA polymerase II, 130kD), an RP117 (60S ribosomal protein L17) pseudogene, a gene for a novel protein similar to Pleurodeles waltlil RAP55 protein, ESTs, STSS and two CpG islands, complete sequence.	
ACCESSION	Al137077	
VERSION	Al137077.31 GI:113445287	
KEYWORDS	HTG; CpG island; RAP55; ribosomal protein; RNA polymerase II; RPL17; TAF2C1; TATA box binding protein; TBP-associated factor. human.	
SOURCE	Homo sapiens	
ORGANISM		

```

repeat_region 5174. .5275
/note="2 copies 51 mer 88% conserved"
repeat_region 5818. .5863
/note="23 copies 2 mer gt 97% conserved"
repeat_region 5920. .5977
/note="29 copies 2 mer tg 89% conserved"
repeat_region 6091. .6148
/note="MIR repeat: matches 99. .158 of consensus"
repeat_region 8161. .8693
/note="MIR2A repeat: matches 6. .453 of consensus"
repeat_region 8775. .9073
/note="AluJo repeat: matches 5. .302 of consensus"
repeat_region 9270. .9587
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 9597. .9766
/note="Alusg/x repeat: matches 129. .302 of consensus"
repeat_region 12354. .12493
/note="MER86 repeat: matches 15. .157 of consensus"
repeat_region 12892. .13263
/note="HAL1 repeat: matches 588. .993 of consensus"
repeat_region 14836. .15072
/note="L1ME2 repeat: matches 4457. .4675 of consensus"
repeat_region 15073. .15364
/note="Alusx repeat: matches 1. .303 of consensus"
repeat_region 15365. .15404
/note="L1ME2 repeat: matches 4675. .4711 of consensus"
repeat_region 15405. .15690
/note="Alusx repeat: matches 27. .312 of consensus"
repeat_region 15691. .15726
/note="L1ME2 repeat: matches 4711. .4743 of consensus"
repeat_region 15727. .15997
/note="Alusx repeat: matches 14. .288 of consensus"
repeat_region 15998. .16499
/note="L1ME2 repeat: matches 4743. .5207 of consensus"
repeat_region 16500. .16797
/note="Alusx repeat: matches 1. .295 of consensus"
repeat_region 16798. .17341
/note="L1ME2 repeat: matches 5207. .5332 of consensus"
repeat_region 17342. .17678
/note="THE1B repeat: matches 1. .363 of consensus"
repeat_region 17679. .17920
/note="L1ME2 repeat: matches 5932. .6155 of consensus"
repeat_region 20266. .20475
/note="MER8 repeat: matches 1. .239 of consensus"
repeat_region 20828. .21125
/note="AluY repeat: matches 1. .298 of consensus"
repeat_region 21127. .21209
/note="L1MC4 repeat: matches 7883. .7976 of consensus"
repeat_region 21400. .21433
/note="17 copies 2 mer ca 88% conserved"
repeat_region 21539. .21837
/note="Alusg repeat: matches 1. .300 of consensus"
repeat_region 25368. .25672
/note="Alusg repeat: matches 1. .305 of consensus"
repeat_region 25674. .25974
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 26175. .26238
/note="L1MC/D repeat: matches 4303. .4365 of consensus"
repeat_region 26239. .26542
/note="AluJo repeat: matches 1. .297 of consensus"
repeat_region 26543. .26736
/note="L1MC/D repeat: matches 4365. .4557 of consensus"
repeat_region 26737. .27032
/note="AluY repeat: matches 1. .296 of consensus"
repeat_region 27033. .27855
/note="L1MC/D repeat: matches 4557. .5389 of consensus"
repeat_region 27933. .28237
/note="Alusg repeat: matches 1. .298 of consensus"
repeat_region 28233. .28360
/note="L1MC/D repeat: matches 5417. .5526 of consensus"
repeat_region 28423. .28698

```

[illegible]

RESULT	14	
AF034801		
LOCUS	1500 bp	PR1
DEFINITION	Human sapiens liprin-alpha4 mRNA, partial cds.	13-JUL-1998
ACCESSION	AF034801	
VERSION	AF034801.1	GI:3309536
KEYWORDS	human.	
SOURCE	Human sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 1500)	
AUTHORS	Serra-Pages,C., Medley,O.G., Tang,M., Hart,A. and Streuli,M.	
TITLE	Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins	
JOURNAL	J. Biol. Chem. 273 (25), 15611-15620 (1998)	
MEDLINE	98288299	
REFERENCE	2 (bases 1 to 1500)	
AUTHORS	Serra-Pages,C. and Streuli,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (18-NOV-1997) Tumor Immunology, Dana-Farber Cancer Institute, 44 Binney St, Boston, MA 02115, USA	
FEATURES	Location/Qualifiers	

source

CDS

1..1500

/organism="Homo sapiens"

/db_xref="taxon.9606"

<_>1500

/codon_start=1

/product="liprin-alpha4"

/protein_id="AAC26102.1"

/db_xref="GI:3309537"

/translation="SIPTSLTSLASAPPLSGRSTPKLTSSRAAOI.DRMGVMTLP

SDLKRRKLLSPVSRRENRDKATIKCTSPSPPTLRLEKLGHPALSGEKGKSL

EDGNSPSSNSQDLSLHKCAKRGKIKSSIGRLFGKKEKGRLOLRDCAITGHVLT

SEFSQKQPMVPAKLGCTQAEKDRBLKKHQLLEDDARRKGMFAQNDGPTVSMLELWG

PMTVAACRANVKRSGAINSDALDTEIQREIGISNALRKLRLCALQSMVSLTSPAP

PTSTSSGNVVYTHEEMTLETSTKTTLAYGDMNHETGKWLPSGLPQPSYPMEC

LVDRLMDLHTKDLRVLHVKMVDSPFRTSLQYIMCLRLNDRKLEKREESQHEI

KDVLVMTNDQVHWVQSIGLRVAGNLHESGVHGALLADENFDNLTALILQIPTON

TOARVNEREFNNLALGCTDKLDDDDKVFRRAPSWRRKRPREHHGRA"

BASE COUNT

352 a 437 c 450 g 261 t

Query Match

Best Local Similarity 66.3%; Score 23.2; DB 9; Length 1500;

Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

8 qqagcaggcagcaccacccagccgagc 35

Db

309 GGATCAGGCGACGACACCCCGAGCAGC 336

RESULT 15

AL451082/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL451082

200794 bp

DNA

HTG

06-JUN-2001

Homo sapiens chromosome 1 clone RP11-307B16, *** SEQUENCING IN

PROGRESS ***, 7 unordered pieces.

AL451082 AC025413

AL451082.3 GI:14330036

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Plumb,B.

1 (bases 1 to 200794)

Direct Submission

Submitted (06-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequest@sanger.ac.uk

On Jun 8, 2001 this sequence version replaced gi:13785420.

Draft Sequence Produced by Whitehead Institute/MIT Center for

Genome Research, 320 Charles Street,

Cambridge, MA 02141, USA

http://www-seq.wi.mit.edu

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: ba307816

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: M13; M7815; 48% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 198149 bases at least Q40

Consensus quality: 198873 bases at least Q30

Consensus quality: 199492 bases at least Q20

Insert size: 200194; sum-of-contigs

Insert size: 206715; 1.9% error; agarose-fp

Quality coverage: 7.55x in Q20 bases; sum-of-contigs Quality

coverage: 7.31x in Q20 bases; agarose-fp

FEATURES

source

1..200794

/organism="Homo sapiens"

/db_xref="taxon.9606"

/chromosome="1"

/clone="RP11-307B16"

/clone_lib="RPC1-11.2"

1..82883

/note="assembly_fragment:00748

clone_end:T7

vector_side:left"

82984..138290

/note="assembly_fragment:01414"

138391..141724

/note="assembly_fragment:02038"

141825..158990

/note="assembly_fragment:00744

fragment_chain:1"

159091..169729

/note="assembly_fragment:00176

fragment_chain:1"

169830..177240

/note="assembly_fragment:01630

fragment_chain:1"

177341..200794

/note="assembly_fragment:01971

fragment_chain:1

vector_side:right"

clone_end:SP6

BASE COUNT

53635 a 50216 c 47323 g 49020 t

ORIGIN

Query Match

Best Local Similarity 66.3%; Score 23.2; DB 2; Length 200794;

Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

8 ggagcagggcagcaccacccagccgagc 35

Db

120544 GGATCAGGCGACGACACCCCGAGCAGC 120517

Search completed: February 20, 2002, 14:26:46

Job time: 14445 sec

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM40718.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 3863; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5381 BP; 1249 A; 1438 C; 1562 G; 1132 T; 0 other:

Query Match 66.3%; Score 23.2; DB 22: Length 5381;
Best Local Similarity 89.3%; Pred. No. 34;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 8 ggaagcaggcagcaccgccagcagc 35
||| ||||| ||||| ||||| |||||
Db 1633 ggaacagggcagcaaccgccagcagc 1660

RESULT 2
AA159875
ID AA159875 standard; cDNA: 5381 BP.
XX
AC AA159875;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3864.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM40719.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 3864; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5381 BP; 1249 A; 1438 C; 1562 G; 1132 T; 0 other:

Query Match 66.3%; Score 23.2; DB 22: Length 5381;
Best Local Similarity 89.3%; Pred. No. 34;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 8 ggaagcaggcagcaccgccagcagc 35
||| ||||| ||||| ||||| |||||
Db 1633 ggaacagggcagcaaccgccagcagc 1660

RESULT 3
AA158089
ID AA158089 standard; cDNA: 5387 BP.
XX
AC AA158089;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 292.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM38933.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1: SEQ ID NO 292; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SO Sequence 5387 BP; 1244 A; 1451 C; 1560 G; 1132 T; 0 other:
 QY 8 ggagcagggcagcaccgccagccagc 35
 Db 1677 ggatcagggcagcaaccgccagcagc 1704
 RESULT 4
 AA158088
 ID AA158088 standard; cDNA: 5414 BP.
 XX
 AC AA158088;
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 291.
 XX
 KW Human: neurotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.

XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM38932.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1: SEQ ID NO 291; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SO Sequence 5414 BP; 1250 A; 1456 C; 1573 G; 1135 T; 0 other:
 QY 8 ggagcagggcagcaccgccagccagc 35
 Db 1677 ggatcagggcagcaaccgccagcagc 1704
 RESULT 5
 AAC75948/c
 ID AAC75948 standard; cDNA: 799 BP.
 XX
 AC AAC75948;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1503 polynucleotide sequence SEQ ID NO:3005.
 XX
 KW Human: open reading frame; ORFX; detection; cytoskeletal; hepatotropic;
 KW vulnereary; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease.

XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
OS Homo sapiens.
XX
XX WO200058473-A2.
PN
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
PI Shinkets RA, Leach M;
DR WPI: 2000-602362/57.
PT P-PSDB: AAB41739.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 2237; 5507pp; English.
PS
CC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397
CC sequences represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticoagulant; antirheumatic; immunosuppressive;
CC immunostimulant; cardiant; thermolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antifungal; antibacterial; antiviral; antifungal; antineuritic;
CC antihypertensive; antinaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 799 BP; 165 A; 250 C; 252 G; 132 T; 0 other;

Query Match	63.4%	Score 22.2;	DB 21;	Length 799;
Best Local Similarity	88.9%;	Pred. No. 74;		
Matches 21; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

OY 1 cccctcgyagcaggcgaccacca 27
 || | | | | | | | | | | |
Db 616 CCACGGCGGAGCAGGGCAGCACCTCCA 590

RESULT	6
AA161340/c	
ID	AA161340 standard; cDNA; 956 BP.
XX	
AC	AA161340;
XX	

DT	22-OCT-2001	(first entry)
XX		
DE	Human polynucleotide SEQ ID NO 5329	

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

OS Homo sapiens

PN W0200153312-A1

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

09-JUL-2000; 2000US-0598042.

03-AUG-2000; 2000US-0653450.

19-OCT-2000; 2000US-0693036.

XX
PR
XX

FA (HISE) HIDE LING
XX

PI Wang J., Wang Z., Wehrman T., Xu C., Xue A.J., Yang Y., Zhang J.;

PI Zhao QA, Zhou P, Goodrich K, Dimmock NJ, YX

DR WP1; 2001-442233/41
DR P-PSDB; AAMA2184

Novel nucleic acids and polynucleotides, useful for treating disorders

such as central nervous system injuries -

PS Claim 1; SEQ ID NO 5329; 100/8pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and

immunosuppressant and cytostatic activity. The polynucleotides are used in a composition containing a polypeptide or polynucleotide.

of the invention may be used to treat diseases of the peripheral nervous system, including peripheral neuropathy.

CC localised neuropathies and central nervous system diseases, such as

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC Note: The sequence data for this patent did not form part of the printed

Spectroscopy

XX
XX
XX

2000

Query Match 63.48; Score 22.2; DB 22; Length 956;

```

best local similarity: 80.00; 100.00; 100.00;
Matches: 24; Conservative: 0; Mismatches: 3; Indels: 0; Gaps: 0;

```

QY 1 cccctcgagagcagggcagcaccacca 27
 11 1 111111111111111 111
 Db 262 CCACGCGGGAGCAGGGCAGCACCTTCCA 236

RESULT 7
AA159554/C

ID AA159554 standard; cDNA: 1525 BP.
AC AA159554:
XX
XX 22-OCT-2001 (first entry)
DT
XX Human polynucleotide SEQ ID NO 1757.
DE
XX Human: nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW Leukemia; ss.
XX
XX Homo sapiens.
OS
XX MO200153312-A1.
PN
XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000MO-US34263.
PE
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
PI
XX WPI: 2001-442253/47.
DR P-PSDB: AAM40398.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
XX Claim 1; SEQ ID NO 1757; 10078pp; English.
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1525 BP; 306 A; 445 C; 509 G; 265 T; 0 other;
SO

Query Match 63.4%; Score 22.2; DB 22; length 1525;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cccctggagagcagcagcagcccca 27
DB 542 CCACCGGAGCAGGCGACGACCTCCA 516
XX

RESULT 8
AAC00584
ID AAC00584 standard; cDNA: 271 BP.
XX
XX AAC00584:
AC
XX 06-OCT-2000 (first entry)
DT
XX Human secreted protein 5' EST, SEQ ID NO: 582.
DE
XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PE
XX 26-FEB-1999; 99US-0122487.
PR (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI: 2000-500381/45.
DR P-PSDB: AAC00578.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS
XX Claim 1; SEQ ID 582; 71pp + CD-ROM; English.
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 271 BP; 49 A; 100 C; 72 G; 50 T; 0 other;
SO

Query Match 62.3%; Score 21.8; DB 21; length 271;
Best Local Similarity 78.8%; Pred. No. 1e+02; 7;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 cctcgagcagcagcagcccccagcagc 35
DB 94 cctcgagcagcagcagccacacacagcagc 126
XX

RESULT 9
AAZ42266
ID AAZ42266 standard; cDNA: 274 BP.
XX
XX AAZ42266:
AC
XX 01-FEB-2000 (first entry)
DT
XX

DE	Human 5 EST isolated from a cDNA library SEQ ID NO:25.
XX	
KW	Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM	gene therapy; chromosome mapping; upstream regulatory sequence;
KW	forensic; location; development; protein synthesis; stability;
KX	regulation; identification; ss.
XX	
OS	Homo sapiens.
PN	MO9953051-A2.
PD	21-OCT-1999.
XX	
PF	09-APR-1999; 99NC-IB00712.
XX	
PR	09-APR-1998; 98US-0057719.
PR	28-APR-1998; 98US-0069047.
XX	
PA	(GEST) GENSET.
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
DR	WPI: 2000-038446/03.
DR	P-PSDB: AAY64652.
XX	
PT	Novel secreted protein 5' expressed sequence tag sequences used in
PT	diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX	
XX	Claim 1; Page 184-185; 837pp; English.
XX	
CC	AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC	sequences, corresponding to human secreted proteins. AAY64651 to
CC	AA243075 represent the EST-related proteins corresponding to AA242265 to
CC	AA243052. The 5' ESTs can be used for producing secreted human gene
CC	products. They can be used to identify and isolate 5' untranslated
CC	regions (UTRs) and upstream regulatory regions which control the
CC	location, development stage, rate, and quantity of protein synthesis, as
CC	well as stability of mRNA. The ESTs are also useful as probes for
CC	chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC	also be used in forensic procedures to identify individuals, or in
CC	diagnostic procedures to identify individuals having genetic diseases
CC	resulting from abnormal gene expression. The products may also be used in
CC	gene therapy protocols. The nucleic acids encoding signal peptides can be
CC	used for directing extracellular secretion of a polypeptide or the
CC	insertion of a polypeptide into a membrane, or importing a polypeptide
CC	into a cell. The proteins encoded by the EST sequences may be useful in
CC	treating a variety of human conditions. Secreted proteins have
CC	therapeutic value, and the identification of new secreted proteins is
CC	valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
CC	sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 274 BP; 49 A; 100 C; 75 G; 50 T; 0 other:
OY	3 cctcggagcaggccacccaccgcccac 35
Db	
	94 cctcgggagcaaggcgccactcacgaccacgc 126
RESULT 10	
ID	AAF25163
XX	AAF25163 standard; cDNA; 351 BP.
AC	AAF25163:
XX	
DT	30-APR-2001 (first entry)
XX	
DE	Nucleotide sequence of a human protein having a hydrophobic domain.

XX	Human: hydrophobic protein; secretory protein; membrane protein; sepsis;	
KW	tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;	
KV	infectious disease; cancer; ulcer; periodontal disease; coagulation;	
KX	Parkinson's disease; fertility; immune response; thrombosis; ss.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	1..351	
FT	CDS	
FT	/*tag= a	
FT	/product= "hydrophobic protein"	
FT	/Note= "no termination codon given"	
XX		
PN	W0200104297-A2.	
PD	18-JAN-2001.	
XX		
PF	16-JUN-2000; 2000MO-JP03942.	
XX		
PR	08-JUL-1999; 99JP-0194359.	
XX		
PA	(SAGA) SAGAMI CHEM RES CENT.	
TA	(PROT-) PROTEGENE INC.	
XX		
PI	Kato S, Kimura T;	
XX		
DR	WPI: 2001-103081/11.	
DR	P-PSDB: AAB31673.	
XX		
PT	Isolated human proteins and polynucleotides are used in research and	
PT	have activities including cell proliferation/differentiation activity,	
PT	immune stimulating activity and receptor/ligand activity -	
XX		
PS	Claim 3; Page 112; 151pp; English.	
XX		
CC	The present sequence encodes a human protein with hydrophobic domains.	
CC	AMF25173 represents a longer version of the present sequence. The	
CC	protein possesses a hydrophobic domain and so is a secretory protein	
CC	or a membrane protein. The protein is used as an antigen to prepare	
CC	antibodies. The polynucleotide sequence is useful as a source of probes	
CC	for genetic diagnosis. It is also useful for producing the protein	
CC	in large quantities and for gene therapy. The eukaryotic cells are used	
CC	for detecting the receptors or ligands corresponding to the protein and	
CC	for detecting small novel pharmaceuticals. The antibodies are also used	
CC	for detection, quantification and purification of the proteins. Both the	
CC	protein and polynucleotide may be used in research or as nutritional	
CC	sources or supplements. The protein may have cytokine and cell	
CC	proliferation/differentiation activity, immune stimulating or suppressing	
CC	activity, hematopoiesis regulating activity, tissue growth activity,	
CC	activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic	
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory	
CC	activity and tumour inhibition activity. It may therefore be used to	
CC	treat immune deficiencies resulting from autoimmune disorders or	
CC	infectious diseases, cancer, sepsis, anaemias, burns and ulcers,	
CC	periodontal disease, Parkinson's disease, induce fertility, improve	
CC	immune response and enhance coagulation or inhibit thrombosis.	
XX		
XX	Sequence 351 BP; 67 A; 136 C; 77 G; 71 T; 0 other;	
XX		
Query Match	62.3%; Score 21.8; DB 22; Length 351;	
Best Local Similarity	78.8%; Pred. No. 1e+02; 7;	
Matches 26; Conservative	0; Mismatches	
	Indels 0; Gaps 0;	
QY	3 cctctggagcagggcagaccaccacgcacacg 35	
Db	60 cctctgggacacagggcgccaccctcagccaccagc 92	
RESULT 11		
ID	AAV86229	
XX	AAV86229 standard; cDNA: 355 BP.	


```

XX Haematopoietic-specific protein; HSP; human; agonist; antagonist;
KM Autoimmune disease; graft rejection; malignancy; cancer; infection;
KM hypersensitivity; therapy; diagnosis; ss.
OS Homo sapiens.
XX Key
XX CDS Location/Qualifiers
FT sig_peptide 42..611 /*tag= a
FT sig_peptide 42..107 /*tag= b
FT mat_peptide 108..608
FT /*tag= c
XX
XX WO9738003-A1.
XX
XX 16-OCT-1997.
XX
XX 11-APR-1996; 96MO-US04930.
XX
XX 11-APR-1996; 96MO-US04930.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Li H, Wei Y;
XX
XX WPI: 1997-512641/47.
XX
XX P-PSDB: AAM35904.
XX
XX New nucleic acid encoding human haematopoietic-specific protein -
PT and agonists and antagonists, used for treatment and diagnosis of
PT autoimmune disease, graft rejection, malignancies, infections and
PT hypersensitivity
XX
XX Claim 8; Page 57; 72pp; English.
XX
XX This DNA sequence includes a claimed coding region for human
CC haematopoietic-specific protein (HSP) (AAM35904), a protein that is
CC expressed specifically in thymus and haematopoietic tissues. It
CC was isolated from a cDNA library produced from human B cell.
CC Lymphoma cells (no more details). Typically, the 833 bp sequence,
CC or bases 108-833 that encode amino acids -22 to 167 or 1 to 167 of
CC HSP, the latter being the mature protein, is amplified from a clone
CC deposited at ATCC 97455, and recombinant HSP is expressed in e.g.
CC E. coli, insect, mammalian or plant host cells. The HSP (or DNA
CC encoding it in gene therapy) can regulate differentiation and
CC maturation of immune system cells and can thus be used to treat and
CC protect against autoimmune disease, graft rejection, malignant cells
CC and infection by viruses, fungi and bacteria. Fragments of HSP
CC nucleic acids are used as diagnostic probes or primers, e.g.
CC to detect mutations, or for chromosome identification. Antisense
CC sequences can be used as antagonists of HSP and used to treat or
CC prevent delayed hypersensitivity.
XX
XX Sequence 833 BP; 200 A; 245 C; 244 G; 144 T; 0 other;

```

```

Query Match 62.3%; Score 21.8; DB 18; Length 833;
Best Local Similarity 78.8%; Pred. No. 1e-02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 3 cctcggagcagcagcagcagcagcagcagc 35
DB 101 cctcggagcagcagcagcagcagcagcagc 133

```

```

RESULT 14
AAC59795
ID AAC59795 standard; DNA: 847 BP.
XX
AC AAC59795;
XX

```

```

DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein encoding DNA clone vq2 1.
XX
XX Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
XX systemic lupus erythematosus; rheumatoid arthritis; anemia; stroke;
XX haematologic lupus regulation; tissue growth; wound healing; haemophilia;
XX Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
XX contraceptive; infection; growth inhibition; hyperproliferative disorder;
XX psoriasis; ds.
XX
XX Homo sapiens.
XX
XX WO200055375-A1.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000MO-US07285.
XX
XX 17-MAR-1999; 99US-0124808.
XX
XX 17-MAR-1999; 99US-0124916.
XX
XX 17-AUG-1999; 99US-0149639.
XX
XX 01-OCT-1999; 99US-0157247.
XX
XX 29-NOV-1999; 99US-0167824.
XX
XX 15-FEB-2000; 2000US-0182711.
XX
XX (ALPH-) ALPHAGENE INC.
XX
XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
XX WPI: 2000-638211/61.
XX
XX P-PSDB: AAB34694.
XX
XX Novel proteins and polypeptides useful for the treatment of e.g
PT multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
PT ulcers -
XX
XX Claim 24; Page 396; 493pp; English.
XX
XX This invention relates to 59 human secreted proteins and the nucleotide
CC sequences encoding them. Sequences AAC59788-C59806 and AAB34687-B34745
CC represent the proteins and their encoding nucleotide sequences, and
CC sequences AAB34746-B34771 represent fragments of the proteins. Probes
CC for the DNA sequences are represented by sequences AAC59847-C59956. The
CC proteins exhibit neuroprotective, dermatological, immunosuppressive,
CC antiinflammatory, antianaemic, nootropic, antiparkinsonian,
CC cerebroprotective, haemostatic, vulnerary, cytoskeletal, antiporiatic,
CC antibacterial, virucide, and fungicide activity. The proteins and
CC nucleotide sequences are useful as nutritional sources or supplements
CC and in research. The proteins are useful for treating immune deficiency
CC and disorders, which may be genetic or resulting from infections,
CC autoimmune disorders such as multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid
CC cell deficiencies such as anaemias by regulating haematopoiesis. The
CC proteins are also useful in compositions for bone, cartilage, tendon,
CC ligament and/or nerve tissue growth or regeneration, for wound healing,
CC tissue repair and replacement and in the treatment of wounds, incisions
CC and ulcers. Other uses include in the treatment of central and
CC peripheral nervous system and neuropathies such as Alzheimer's and
CC Parkinson's diseases and Shy-Drager syndrome, and mechanical and
CC traumatic disorders, such as spinal cord disorders, head trauma and
CC stroke. The proteins may also be used as a contraceptive, and for
CC treating coagulation disorders such as haemophilia. The protein and
CC nucleotide sequences with cadherin activity are useful for treating
CC cancer. Other uses for the protein include for inhibiting the growth,
CC infection or function of, or killing, infectious agents such as bacteria,
CC virus, fungi and other parasites, for effecting bodily characteristics
CC such as height, weight, hair colour, effecting biorhythms or cardiac
CC cycles or rhythms, effecting metabolism, catabolism, anabolism,
CC processing, utilization, storage or elimination of dietary fat, lipid,
CC protein, carbohydrate, vitamins, minerals, cofactors, effecting
CC behavioural characteristics, providing analgesic effects and for treating

```

CC hyperproliferative disorders such as psoriasis.
 XX Sequence 847 BP; 202 A; 252 C; 247 G; 146 T; 0 other;
 SQ

Query Match 62.3%; Score 21.8; DB 21; Length 847;
 Best Local Similarity 78.8%; Pred. No. 1e+02;
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 cctcggagacagggcagaccgccagccgacg 35
 ||||||| ||||||| ||| | ||||||| |||||
 DB 117 cctcggagacagggcagccactcagccacgacg 149

RESULT 15
 AAC77973
 ID AAC77973 standard; cDNA; 898 BP.

AC AAC77973:

DT 08-FEB-2001 (first entry)

DE Human cancer associated gene sequence SEQ ID NO:367.

XX Human: cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antistatic; antirheumatic; antiallergic; antiviral;
 KW antineoplastic; antihypertensive; thrombolytic; coagulant; neurotropic;
 KW dermatological; neuroprotective; antidiabetic; antihypertensive;
 KW vasotropic; antiproliferative; antineoplastic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.

OS Homo sapiens.

XX W0200055350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-58753/55.

XX P-PSDB: AAB43764.

XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -

XX Claim 1; Page 915; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnery; immunomodulator;
 CC antidiabetic; antistatic; antirheumatic; antiallergic; antiviral;
 CC antineoplastic; antihypertensive; thrombolytic; coagulant;
 CC dermatological; neuroprotective; antidiabetic; antihypertensive;
 CC vasotropic; antiproliferative; antineoplastic; gene therapy; inflammation;
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate

CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies, to
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 898 BP; 215 A; 258 C; 269 G; 152 T; 4 other;
 SQ

Query Match 62.3%; Score 21.8; DB 21; Length 898;
 Best Local Similarity 78.8%; Pred. No. 1e+02;
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 cctcggagacagggcagaccgccagccgacg 35
 ||||||| ||||||| ||| | ||||||| |||||
 DB 154 cctcggagacagggcagccactcagccacgacg 186

Search completed: February 20, 2002, 14:31:04
 Job time: 12453 sec

Thu Feb 21 07:25:49 2002

us-09-904-420a-1_copy_882_916.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:32:31 : Search time 162.84 Seconds
(without alignments)
48.678 Million cell updates/sec

Title: US-09-904-420A-1_COPY_882_916
Perfect score: 35
Sequence: 1 cccctgggagcaggcagcagcccccagcgcagc 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.8	62.3	833	2	US-08-837-029-1
2	21.6	61.7	1635	3	US-09-234-332-4
3	21.2	60.6	2308	4	US-09-382-256-9
4	21.2	60.6	2308	4	US-09-395-115-9
5	21	60.0	152	1	US-08-136-277-9
6	21	60.0	152	2	US-08-479-403-9
7	21	60.0	152	3	US-08-835-734-9
8	21	60.0	702	1	US-08-136-277-20
9	21	60.0	702	2	US-08-479-403-20
10	21	60.0	702	3	US-08-835-734-20
11	21	60.0	1168	2	US-08-530-569B-11
12	21	60.0	1235	2	US-08-530-569B-12
13	21	60.0	1311	2	US-08-530-569B-1
14	20.8	59.4	2183	4	US-08-808-346-3
15	20.6	58.9	591	3	US-08-833-488B-6
16	20.6	58.9	591	3	US-08-833-488B-7
17	20.6	58.9	591	3	US-08-833-488B-11
18	20.6	58.9	591	3	US-08-833-488B-12
19	20.6	58.9	591	3	US-08-833-488B-27
20	20.6	58.9	591	3	US-08-833-488B-29
21	20.6	58.9	597	3	US-08-833-488B-16
22	20.6	58.9	597	3	US-08-833-488B-17
23	20.6	58.9	609	3	US-08-833-488B-3
24	20.6	58.9	609	3	US-08-833-488B-5
25	20.6	58.9	609	3	US-08-833-488B-8
26	20.6	58.9	609	3	US-08-833-488B-10
27	20.6	58.9	617	3	US-08-833-488B-13

28	20.6	58.9	617	3	US-08-833-488B-15
29	20.6	58.9	759	3	US-08-833-488B-22
30	20.6	58.9	759	3	US-08-833-488B-23
31	20.6	58.9	991	3	US-08-833-488B-19
32	20.6	58.9	991	3	US-08-833-488B-21
33	20	57.1	1504	2	US-08-850-910A-17
34	20	57.1	1788	1	US-08-035-928-1
35	19.8	56.6	190	2	US-08-890-980-18
36	19.8	56.6	190	3	US-08-890-979-18
37	19.8	56.6	190	3	US-09-032-894-18
38	19.8	56.6	190	4	US-09-031-626-18
39	19.8	56.6	1002	2	US-08-890-980-5
40	19.8	56.6	1002	3	US-08-890-979-5
41	19.8	56.6	1002	3	US-09-032-894-5
42	19.8	56.6	1002	3	US-09-032-894-95
43	19.8	56.6	1002	4	US-09-031-626-5
44	19.8	56.6	1002	4	US-09-031-626-95
45	19.8	56.6	3441	2	US-08-525-864A-1

ALIGNMENTS

RESULT 1
US-08-837-029-1
: Sequence 1, Application US/08837029
: Patent No. 5945303
: GENERAL INFORMATION:
: APPLICANT: Wei et al.
: TITLE OF INVENTION: Human Hematopoietic - Specific Protein

: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: MD
: COUNTRY: US
: ZIP: 20850

: COMPUTER READABLE FORM:
: MEDIUM TYPE: IBM PC compatible
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/08/837,029
: APPLICATION NUMBER: WO PCT/US96/04930
: FILING DATE: 11-APR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 11-APR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PF268
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 301-309-8504
: TELEFAX: 301-309-8439
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 833 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 42..608
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 42..107
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 108..608

Sequence 15, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 17, Appl
Sequence 1, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 95, Appl
Sequence 5, Appl
Sequence 95, Appl
Sequence 1, Appl

```

RESULT      3
US-09-382-256-9/c
: Sequence 9, Application US/09382256A
: Patent NO. 6207814
: GENERAL INFORMATION:
: APPLICANT: MIYAZONO, Kohei
: TEN DIJKE, Peter
: FRANZEN, Petra
: YAMASHITA, Hidetoshi
: HELDIN, Carl-Henrik
: TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS,
: HAVING SERINE THREONINE KINASE DOMAINS,
: AND THEIR USE
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fulbright & Jaworski L.L.P.
: STREET: 666 Fifth Avenue
: CITY: New York City
: STATE: New York
: COUNTRY: USA
: ZIP: 10103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS

```

Query Match 60.0%; Score 21; DB 1; Length 152;
Best Local Similarity 82.8%; Pred. No. 47;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

RECORD 6
US-08 479-403-9/c
: Sequence 9, Application US/08479403
: Patent No. 5869039
: GENERAL INFORMATION:
: APPLICANT: MANDEL, Jean-Louis
: APPLICANT: AUBOURG, Patrick
: APPLICANT: MOSSER, Jean
: APPLICANT: SARDE, Claude
: TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Young & Thompson
: STREET: 745 South 23rd Street
: CITY: Arlington
: STATE: VA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/08/479,403

APPLICATION NUMBER: US/08/479,403

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: PATCH, Andrew J.

REGISTRATION NUMBER: 32,925

REFERENCE/DOCKET NUMBER: B2272DIV

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-521-2297

TELEFAX: 703-685-0573

TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 152 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-479-403-9

Query Match

Best Local Similarity 60.0%; Score 21; DB 2; Length 152;

Mismatches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 gggagcaggcagcccccagccgacgc 35

|| ||||| ||||| ||||| |||||

DB 68 GGCAGCAGGTCAGCACCTGCAGCAGCAGC 40

RESULT 7

US-08-835-734-9/c

; Sequence 9, Application US/08835734

; Patent No. 6013769

; GENERAL INFORMATION:

; APPLICANT: MANDEL, Jean-Louis

; APPLICANT: AUBOURG, Patrick

; APPLICANT: MOSSER, Jean

; APPLICANT: SARDE, Claude

; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND

; TITLE OF INVENTION: CORRESPONDING PROTEIN

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Young & Thompson

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/479,403

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/479,403

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: B2272DIV

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; TELEX: 248425 EMBON

; INFORMATION FOR SEQ ID NO: 9:

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/08/479,403

APPLICATION NUMBER: US/08/479,403

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: PATCH, Andrew J.

REGISTRATION NUMBER: 32,925

REFERENCE/DOCKET NUMBER: B2272DIV

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-521-2297

TELEFAX: 703-685-0573

TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 152 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-479-403-9

Query Match

Best Local Similarity 60.0%; Score 21; DB 2; Length 152;

Mismatches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 gggagcaggcagcccccagccgacgc 35

|| ||||| ||||| ||||| |||||

DB 68 GGCAGCAGGTCAGCACCTGCAGCAGCAGC 40

RESULT 7

US-08-835-734-9/c

; Sequence 9, Application US/08835734

; Patent No. 6013769

; GENERAL INFORMATION:

; APPLICANT: MANDEL, Jean-Louis

; APPLICANT: AUBOURG, Patrick

; APPLICANT: MOSSER, Jean

; APPLICANT: SARDE, Claude

; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND

; TITLE OF INVENTION: CORRESPONDING PROTEIN

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Young & Thompson

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/479,403

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/479,403

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: B2272DIV

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; TELEX: 248425 EMBON

; INFORMATION FOR SEQ ID NO: 9:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-08-530-5698-12

RESULT 14
US-08-808-346-3/c
; Sequence 3, Application US/08808346
; Patent No. 6251671
; GENERAL INFORMATION:

US-08-808-346-3/c
; Sequence 3, Application US/08808346
; Patent No. 6251671
; GENERAL INFORMATION:

Search completed: February 20, 2002, 14:32:32
Job time: 14411 sec

```

; APPLICANT: Hogan, Brigit L.M.
; TITLE OF INVENTION: Compositions and Methods of Making
; TITLE OF INVENTION: Embryonic Stem Cells
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,346
; FILING DATE: 28-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9823-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-567-2020
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-808-346-3

Query Match 59.44: Score 20.8; DB 4; Length 2183;
Best Local Similarity 78.14; Pred. No. 53;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0:

QY 4 CTCGGGAGCAGGGCAGCACCCCGCCGCGCAGC 35
      ||||| ||||| || ||||| |||||
DB 367 CTCGGGATCGGGCGCGCCCGCCGCGACCCCGCAGC 336

RESULT 15
US-08-833-488B-6/c
; Sequence 6, Application US/08833488B
; Patent No. 6060326
; GENERAL INFORMATION:
; APPLICANT: Frank, Glenn R.
; APPLICANT: Rushlow, Keith E.
; TITLE OF INVENTION: Method to Detect Canine IgE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,488B

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 13:52:34 ; Search time 3793.25 Seconds
(without alignments)
99.150 Million cell updates/sec

Title: US-09-904-420A-1_COPY_882_916
Perfect score: 35
Sequence: 1 ccctcggagcagggcagccacccccagccgcagc 35

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	ID	Description
1	35	100.0	273	10	AA311787 EST182503
2	26.6	76.0	400	11	BG555601 ia84a02.x
3	26.4	75.4	353	11	BG386035 602455204
4	25.4	72.6	229	10	AW872092 db22c01.y
5	25.4	72.6	312	10	AU007753 AU007753
6	25.4	72.6	639	11	B1349635 dae61e11
7	25	71.4	649	11	B1154476 602904810
8	24.2	69.1	573	10	AL589082 AL589082
9	24.2	69.1	587	10	BG712238 pg11n.pk0
10	23.4	66.9	67	10	AI268359 qm05a09.x
11	23.4	66.9	497	10	AA155158 mr98b01.r
12	23.2	66.3	457	13	AQ877378 HS_2145_B

13	23.2	66.3	462	10	AI124575
14	22.8	65.1	497	11	B1403328
15	22.8	65.1	562	10	AI217274
16	22.8	65.1	640	11	B1067375
17	22.8	65.1	662	13	AZ850363
18	22.8	65.1	1326	11	BF303935
19	22.6	64.6	145	11	BF944451
20	22.6	64.6	178	11	BG002207
21	22.6	64.6	189	11	BF955325
22	22.6	64.6	346	10	AJ284096
23	22.6	64.6	354	11	BF953417
24	22.6	64.6	384	11	BE934586
25	22.6	64.6	411	11	BF846035
26	22.6	64.6	542	11	BG014392
27	22.6	64.6	981	10	AL541947
28	22.4	64.0	239	13	CNS03VF4
29	22.4	64.0	632	11	B1390398
30	22.2	63.4	355	10	AI43403
31	22.2	63.4	568	11	B1340223
32	22.2	63.4	744	11	BG715648
33	22.2	63.4	829	13	CNS03Y7H
34	22.2	62.9	228	13	AQ846434
35	22.2	62.9	251	10	AA973990
36	22.2	62.9	375	10	AV630095
37	22.2	62.9	448	10	AA935860
38	22.2	62.9	488	11	B1164168
39	22.2	62.9	492	10	AA988946
40	22.2	62.9	519	10	BE050797
41	22.2	62.9	923	11	BG029934
42	21.8	62.3	235	11	W24845
43	21.8	62.3	256	10	AW404274
44	21.8	62.3	264	10	BB584341
45	21.8	62.3	394	10	BE246824

ALIGNMENTS

RESULT 1

AA311787 273 bp mRNA EST 19-APR-1997
LOCUS EST182503 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
DEFINITION similar to T-cell factor 1, A/B/C, mRNA sequence.
AA311787
ACCESSION AA311787.1 GI:1964114
VERSION EST.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 273)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,K.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-ai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagan,N.S., Glodek,A.,
Gneltm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Yi,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Olsen,H., Raymond,L.,
Kunsch,C., Hungjtin,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venier,J.C.

TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 9606280
COMMENT Contact: Kerlavage, AR

```

/db_xref="taxon:10090"
/clone_lib="Melton Mouse El6 5 Pancreas Library M1621"
/sex="Both"
/tissue_type="Total pancreas"
/dev_stage="Embryonic day 16.5"
/lab_host="TOP10"
/notes="Pancreas; Vector: pZero-2; Site.1: NotI; Site.2: XhoI; Library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. XhoI site destroyed during cloning. Size-selected by column fractionation; average insert size 1.5 kb. Primary library, unamplified."
```

BASE COUNT	87 a	148 c	98 g	67 t
ORIGIN				
Query Match		76.0%	Score 26.6;	DB 11: Length 400;
Best Local Similarity		87.9%	Prod. NO. 75;	
Matches	29;	Conservative	0;	Mismatches 4; Indels 0; Gaps 0;
QY	3	ctctgggagcaggggcagcacccccccagccgcagc	35	
	171	CTCTGGGAGCAGGGCAGCATCCGCAGCCTCAAC	203	

RESULT	3
BG386035	
LOCUS	353 bp mRNA
DEFINITION	602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5', EST 12-MAR-2001
ACCESSION	BG386035
VERSION	BG386035
KEYWORDS	GI:13279481
SOURCE	human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 353)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC

Tissue Procurement: ATCC
 CDNA Library Preparation: Tling Hong/Rubin Laboratory
 CDNA Library Analysis by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone Distribution: WGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LLMC1307 row: h column: 02
 High quality sequence stop: 276.

FEATURES	SOURCE
1. .353	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:4583473"	
/clone_lib="NIH_MGC_15"	
/tissue_type="adenocarcinoma cell line"	
/lab_host="DH10B (phage-resistant)"	
/note="organ: colon; Vector: pOTR7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming, Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"	
80 a	111 c 122 q 40 t
BASE COUNT	

Query Match 75.48; Score 26.4; DB 11; Length 353;

Query Match	Best Local Similarity	Score	DB	Length
72.6%	82.9%	25.4	DB 10	229
Matches	29	Conservative	0	Mismatches 6; Indels 0; Gaps 0;
QY	1	ccctcgggagcagcagcccccagcgcagc 35		
Db	170	CTCCAGGTGACGAGGCGAGCACCAGGCGGCGC 136		
RESULT	5			
LOCUS	AU007753/c	312 bp	mRNA	EST 31-JUL-1998
DEFINITION	AU007753 Schizosaccharomyces pombe late log phase cDNA			
ACCESSION	AU007753			
VERSION	AU007753.1	GI:3344211		
KEYWORDS	EST.			
SOURCE	fission yeast.			
ORGANISM	Schizosaccharomyces pombe			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.			
AUTHORS	1 (bases 1 to 312)			
TITLE	Morimyo, M. and Mita, K.			
JOURNAL	Identification of expressed sequence tags of Schizosaccharomyces pombe			
COMMENT	Unpublished (1998)			
FEATURES	source			
ORIGIN	1..312			
LOCUS	/organism="Schizosaccharomyces pombe"			
DEFINITION	/strain="972"			
ACCESSION	/db_xref="taxon:4896"			
VERSION	/clone="spc02459"			
KEYWORDS	/clone_lib="Schizosaccharomyces pombe late log phase cDNA"			
SOURCE	/sex="h minus"			
ORGANISM	/note="vector: M13mp19; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"			
REFERENCE	BASE COUNT 67 a 71 c 86 g 88 t			
AUTHORS	Query Match	72.6%	Score 25.4	DB 10; Length 312;
TITLE	Best Local Similarity	82.9%	Pred. No. 1.7e+02;	
JOURNAL	Matches	29	Conservative	0; Mismatches 6; Indels 0; Gaps 0;
COMMENT	QY	1	ccctcgggagcagcagcccccagcgcagc 35	
FEATURES	Db	223	CCTCTCAGCAGGCGAGCACCAGGCGGCGC 189	
ORIGIN	RESULT	6		
LOCUS	B1349635/c	639 bp	mRNA	EST 31-JUL-2001
DEFINITION	B1349635			
ACCESSION	B1349635			
VERSION	B1349635.1	GI:15044081		
KEYWORDS	EST.			
SOURCE	African clawed frog.			
ORGANISM	Xenopus laevis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.			
AUTHORS	1 (bases 1 to 639)			
TITLE	daei1e11.y3 Blackshear/Soares normalized Xenopus egg library			
JOURNAL	Xenopus laevis cDNA clone IMAGE:4678748 5' similar to TR:09QXJ4			
COMMENT	O9QX74 ADP-RIBOSYLATION FACTOR-LIKE MEMBRANE-ASSOCIATED PROTEIN.			
FEATURES	BASE COUNT 56 a 56 c 72 g 45 t			
ORIGIN	1..229			
LOCUS	/organism="Xenopus laevis"			
DEFINITION	/db_xref="taxon:8355"			
ACCESSION	/clone="XENOPUS_SOURCE_ID:"			
VERSION	/clone_lib="Xenopus laevis oocyte"			
KEYWORDS	/tissue_type="oocyte (stages 5 and 6)"			
SOURCE	/lab_host="top-10 F."			
ORGANISM	/note="vector: pBluescript SK-; Site_1; EcoRI; Site_2; XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-libraries phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995. Genes Dev. 9, 2923-2935). Normalized by Jihwan Song (Song, Cho and Blumberg, unpublished). Note: This is a Xenopus Gene Collection (XGC) library."			
REFERENCE	BASE COUNT 56 a 56 c 72 g 45 t			
AUTHORS	Query Match	96.4%	Pred. No. 86;	
TITLE	Best Local Similarity	96.4%	Pred. No. 86;	
JOURNAL	Matches	27	Conservative	0; Mismatches 0; Indels 1; Gaps 0;
COMMENT	QY	1	ccctcgggagcagcagcccccagcgcagc 28	
FEATURES	Db	255	CCCTCGGGAGCGGCGAGCACCAGCAGCAG 282	
ORIGIN	RESULT	4		
LOCUS	AW872092/c	229 bp	mRNA	EST 22-JUN-2000
DEFINITION	AW872092			
ACCESSION	AW872092			
VERSION	AW872092.1	GI:8006145		
KEYWORDS	EST.			
SOURCE	African clawed frog.			
ORGANISM	Xenopus laevis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.			
AUTHORS	1 (bases 1 to 229)			
TITLE	Clifton, S., Johnson, S.L., Blumberg, B., Song, J.J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.			

Thu Feb 21 07:25:52 2002

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1095 row: d column: 03
High quality sequence stop: 647.
Location/Qualifiers

FEATURES
source

1..649
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5034194"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NCI_CGAP Library."
157 a 204 c 167 g 121 t

BASE COUNT
ORIGIN

Query Match 71.4%; Score 25; DB 11; Length 649;
Best Local Similarity 84.8%; Pred. No. 2.3e+02;
Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 cctcggagcaggcagcagcccccagccgacg 35
||||| ||||||| ||||||| ||||||| |||
Db 359 CCTCGGAGCAGGCGAGCATCCGCGAGCCTCAAC 391

RESULT 8
AL589082 573 bp mRNA EST 02-MAR-2001
AL589082 BP Chicken Brain Library Gallus gallus cDNA clone
LOCUS ROS083B09, mRNA sequence.
DEFINITION
ACCESSION AL589082
VERSION AL589082.1 GI:13194116
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 573)
Murray, F.
BP Chicken Brain Library
Unpublished (2001)
Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
CGCGCGCGCTTTT TTTT TTTT TTTT TTTT 3' Poly A RNA purchased from Clontech
(*6854-)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

1..573
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="ROS083B09"
/clone_lib="BP Chicken Brain Library"

REFERENCE
AUTHORS

Xenopodinae: Xenopus.
1 (bases 1 to 639)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Powers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R., and Wilson, R., 1999
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library constructed by Bento Soares and M. Fatima Bonaldo (University of Iowa). DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
High quality sequence stop: 432.
Location/Qualifiers

TITLE
JOURNAL
COMMENT

1..639
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:468748"
/clone_lib="Blackshear/Soares normalized Xenopus egg library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI; polyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G., and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery'. Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-drl8 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 X 10⁵ recombinants, with average insert sizes of 1-1.5 kb."
173 a 143 c 176 g 147 t

BASE COUNT
ORIGIN

Query Match 72.6%; Score 25.4; DB 11; Length 639;
Best Local Similarity 82.9%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cccctcggagcaggcagcagcccccagccgacg 35
||||| ||||||| ||||||| ||||||| |||
Db 155 CTCAGGTGAGCAGGCGAGCAGCCCCCGAGCGCAGC 121

REFERENCE
AUTHORS

B1154476 649 bp mRNA EST 05-JUL-2001
B1154476 602904810F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5034194 5', mRNA sequence.
DEFINITION
ACCESSION B1154476
VERSION B1154476.1 GI:14614477
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 649)
NIH-MGC <http://mgi.mgi.nhl.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE
JOURNAL
COMMENT

1 (bases 1 to 649)
NIH-MGC <http://mgi.mgi.nhl.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelsing, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
JOURNAL
COMMENT
The WashU-RHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-RHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LInL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:370881
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 264.

FEATURES
source

1. .497
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:605449"
/clone_lib="Stratagene mouse embryonic carcinoma (#937317)"
/tissue_type="carcinoma"
/dev_stage="embryonic"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. p19 cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."
BASE COUNT 149 a 143 c 125 g 80 t
ORIGIN

Query Match 66.9%; Score 23.4; DB 10; Length 497;
Best Local Similarity 81.8%; Pred. No. 7.1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cctcggagcagcggcagcagcccccagccgacg 34
DB 33 CCCCCAGCGCAGAGCAGACCCCGCAGCGAC 65

RESULT 12
AQ877378

LOCUS AQ877378 457 bp DNA GSS 09-NOV-1999
DEFINITION HS_2145_B2_H07_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2145 Col=14 Row=P, DNA sequence.

ACCESSION AQ877378
VERSION AQ877378.1 GI:6308845
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 457)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2145 row: P column: 14
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 457.

FEATURES
source

1. .457
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-2145 Col=14 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 96 a 145 c 103 g 111 t 2 others
ORIGIN

Query Match 66.3%; Score 23.2; DB 13; Length 457;
Best Local Similarity 89.3%; Pred. No. 8.1e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 ggagcagggcagcagcccccagccgacg 35
DB 321 GGATCAGGCGACACCCCGCAGCGACG 348

RESULT 13
A1124575

LOCUS A1124575 462 bp mRNA EST 11-SEP-1998
DEFINITION am59a03.x1 Johnston frontal cortex Homo sapiens cDNA clone IMAGE:1539820 3' similar to TR:Q13135 Q13135 LAR-INTERACTING PROTEIN 1A. [1] ; mRNA sequence.

ACCESSION A1124575
VERSION A1124575.1 GI:3593089
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 462)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Thelsing, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LInL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 427.

FEATURES
source

1. .462
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1539820"
/clone_lib="Johnston frontal cortex"
/sex="male"
/tissue_type="pooled frontal lobe"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: Bluescript SK-; Site_1: EcoRI
; Stanley Neuropathology Consortium (www.stanleylab.org)
brains S-58, S-65, S-67, S-78. Random + oligo-dT primed into EcoRI site of ZAP II Vector. Mass excised. Avg

Thu Feb 21 07:25:52 2002

us-09-904-420a-1_copy_882_916.rst

Insert length 1.9kb. Custom library provided by Dr. Nancy Johnston [(410) 614-3918, njewelchlink.welch.jhu.edu].

BASE COUNT 118 a 141 c 134 g 69 t
ORIGIN
Query Match 66.3%; Score 23.2; DB 10; Length 462;
Best Local Similarity 89.3%; Pred. No. 8.1e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 ggaagcaggcagcaccacccagccgagc 35
Db 274 GGATCAGGCGAGCAACCCAGCAGCAGC 301

RESULT 14
LOCUS BI403328/c
DEFINITION MI-P-CPI-nwi-c-10-0-UI s1 MI-P-CPI Sus scrofa cDNA clone
ACCESSION BI403328
VERSION BI403328.1 GI:15182389
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE 1 (bases 1 to 497)
JOURNAL Bonaudo,M.F., Lennon,G. and Soares,M.B.
MEDLINE Normalization and subtraction: two approaches to facilitate gene
COMMENT discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Tuggle CK
Iowa State University
201 Kildeer Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized uterus library cDNA Library Preparation: M.B. Soares Lab
, University of Iowa EST sequencing: M.B. Soares Lab, University of
Iowa Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES source
1..497 Location/Qualifiers
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-CPI-nwi-c-10-0-UI"
/clone_lib="MI-P-CPI"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-CPI
library is normalized library derived from the MI-P-CPI
library, ultimately derived from uterus tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigest.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaudo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_LIB=MI-P-CPI
TAG_ISSUE=uterus
TAG_SEQ=AGTCAATCG*
BASE COUNT 63 a 180 c 157 g 97 t

ORIGIN

Query Match 65.1%; Score 22.8; DB 11; Length 497;
Best Local Similarity 79.4%; Pred. No. 1.1e+03;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ccctcggagcaggcagccaccccgccgagc 34
Db 432 CTCCTCGGAGCAGCAGCAGCAGCAGCAGC 399

RESULT 15
LOCUS AI217274/c
DEFINITION QI25401.x1 Soares,NFL,T.GRC.S1 Homo sapiens cDNA clone
IMAGE:1845672.3' similar to gb:J05392 SYNDECAN-1 PRECURSOR (HUMAN
);, mRNA sequence.
ACCESSION AI217274
VERSION AI217274.1 GI:3797089
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 562)
JOURNAL NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicqap.
MEDLINE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: ccapbs-remail.nih.gov

This clone is available royalty-free through LILN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1019 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 255.
Location/Qualifiers
1..562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1845672"
/clone_lib="Soares_NFL_T_GRC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung Nhl19W, testis NHT, and B-cell
NCL-CCAP-CCBI) were mixed and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.E. clones 297480-302087, 642612-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaudo."

BASE COUNT 111 a 126 c 190 g 135 t
ORIGIN

Query Match 65.1%; Score 22.8; DB 10; Length 562;
Best Local Similarity 79.4%; Pred. No. 1.1e+03;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ccctcggagcaggcagccaccccgccgagc 34
Db 363 CCCTCGGAGCAGTAGAGCACCCCTGACGAG 330

Search completed: February 20, 2002, 13:52:37
Job time: 16226 sec

us-09-904-420a-1_copy_882_916.rst

Thu Feb 21 07:25:52 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:26:01 ; Search time 1729.86 Seconds
(without alignments)
333.785 Million cell updates/sec

Title: US-09-904-420A-1_COPY_866_900
Perfect score: 35
Sequence: 1 tcatgattaccacccccctcggagcaggcag 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vi.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	35	100.0	1115	9	HSTCFID	Z47361 H.sapiens T
2	35	100.0	1165	9	HSTCFIE	Z47362 H.sapiens T
3	35	100.0	1254	9	HSTCFIA	X59869 Human TCF-1
4	35	100.0	2814	9	HSTCF1B	X59870 Human TCF-1
5	35	100.0	2855	9	HSTCF1G	X63901 Homo sapien
6	35	100.0	2910	9	HSTCF1C	X59871 Human TCF-1
7	35	100.0	84544	2	AC009012	AC009012 Homo sapi
8	35	100.0	86914	2	AC011336	AC011336 Homo sapi
9	35	100.0	159420	2	AC009017	AC009017 Homo sapi
10	35	100.0	200831	2	AC008608	AC008608 Homo sapi
11	28.6	81.7	1790	10	MMTCSTF	X61385 Mouse mRNA
12	27	77.1	162496	2	AC044846	AC044846 Mus muscu
13	21.2	60.6	318	14	AF172532	AF172532 Hepatitis
14	21.2	60.6	2321	9	HSVDAC2	AF172533 Hepatitis
15	21.2	60.6	79305	9	AC005200	AF151093 Homo sapi
16	21.2	60.6	152838	2	AC011589	AC005200 Homo sapi
17	21.2	60.6	157541	2	AC008523	AC011589 Homo sapi
18	21.2	60.6	200831	2	AC008608	AC008523 Homo sapi
19	21.2	60.6	247196	2	AC073822	AC008608 Homo sapi
20	21.2	60.6	253389	2	AC008605	AC073822 Mus muscu
21	21.2	60.0	112209	10	AC022298	AC008605 Homo sapi
22	21	60.0	169278	2	AC016931	AC022298 Mus muscu
23	21	60.0	214780	2	AC069075	AC016931 Homo sapi
24	21	60.0	446	11	G46174	AC069075 Mus muscu
25	20.8	59.4	628	8	AF055762	G46174 Z7065_1 Zeb
26	20.8	59.4	73667	2	AC021402	AF055762 Rutilus haw
27	20.8	59.4	124946	2	AC022231	AC021402 Homo sapi
28	20.8	59.4	124946	2	AC015669	AL139412 Homo sapi
29	20.8	59.4	158134	2	AC015669	AL139412 Homo sapi
30	20.8	59.4	163452	2	AL139412	AL139412 Homo sapi
31	20.8	59.4	171386	2	AL139130	AL139130 Homo sapi
32	20.8	59.4	171386	2	AC016866	AC016866 Homo sapi
33	20.8	59.4	174572	2	AC064803	AC016866 Homo sapi
34	20.8	59.4	214005	2	AC064803	AC064803 Mus muscu
35	20.8	59.4	218673	2	AL1591390	AL1591390 Mus muscu
36	20.6	58.9	1450	1	AF277219	AF277219 Nocardia
37	20.6	58.9	1457	1	AF277211	AF277211 Nocardia
38	20.6	58.9	1462	1	AF277204	AF277204 Nocardia
39	20.6	58.9	13276	8	AF288043	AF277204 Nocardia
40	20.6	58.9	156772	8	AC073405	AF288043 Cucumits S
41	20.6	58.9	159135	2	AC091039	AC073405 Oryza sat
42	20.6	58.9	170218	2	AC068125	AC091039 Homo sapi
43	20.6	58.9	187802	2	AP003065	AP068125 Homo sapi
44	20.6	58.9	191654	2	AC023176	AP003065 Homo sapi
45	20.6	58.9	197022	2	AC055829	AC023176 Homo sapi
45	20.6	58.9	197022	2	AC055829	AC055829 Homo sapi

ALIGNMENTS

RESULT	1
HSTCFID	HSTCFID
LOCUS	H sapiens TCF-1 mRNA for T cell factor 1 splice form D.
DEFINITION	1115 bp mRNA
ACCESSION	Z47361.1
VERSION	GI:619881
KEYWORDS	splice form D; T cell factor 1.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Enkaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1115)
REFERENCE	Mayer, K., Wolff, E., Clevers, H. and Balhausen, W.G.
AUTHORS	The human high mobility group (HMG)-box transcription factor TCF-1:
TITLE	novel isoforms due to alternative splicing and usage of a new exon
REFERENCE	IXA
AUTHORS	Unpublished
TITLE	2 (bases 1 to 1115)
REFERENCE	van de Wetering, M., Oosterwegel, M., Holsteg, F., Dooyes, D.,
AUTHORS	Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.

Thu Feb 21 07:25:44 2002

```

The Netherlands
2 (bases 1 to 1254)
van de Wetering, M., Oosterwegel, M., Dooljes, D. and Clevers, H.
IDENTIFICATION AND CLONING OF TCF-1, A T LYMPHOCYTE-SPECIFIC
TRANSCRIPTION FACTOR CONTAINING A SEQUENCE-SPECIFIC HMG BOX
EMBO J. 10 (1), 123-132 (1991)
JOURNAL
MEDLINE 91114695
AUTHORS See also X59869-X59871.
TITLE Location/Qualifiers
FEATURES
SOURCE
1. .2814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat and HPB-ALL"
/clone_lib="cDNA"
1. .2814
/gene="TCF-1"
/evidence="experimental"
1. .2814
/gene="TCF-1"
80. .889
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1 splice form B"
/protein_id="CAA42527.1"
/db_xref="GI:36788"
/db_xref="SWISS-PROT:P36402"
/translation="MYKTVYSAFNLMLHYPPSGAGHQPQPPPLHKANOPPHGVPO
LSLYEHNSPHPTAPADISQKVHRLQTPDLGFSYLTSGMGLPHTVSWFTHPS
LMLGSGVPGHPAAIPHPATVPSPGKQLPFDRLKTOAESKAKEAKKPTIKKPLNA
FMLYMKEMRAKVIAGCTLKESAINQILGRWHALSREEQAKYVELARKERQLHMLY
PCWSARDNYGKKRRSRKREKHQESTTGNPRELKGNGQESLSMSSSSPA"
misc_feature 539. .769
/gene="TCF-1"
/feature="HMG box"
810
misc_feature
/gene="TCF-1"
/feature="alternative splice site"
BASE COUNT 320 a 391 c 335 g 208 t
ORIGIN
Query Match 100.0%; Score 35; DB 9; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcatgcattaccacccctcgaggcagcaggcag 35
|||||
Db 117 TCATGATTACCCACCCCTCGGAGCAGGGCAG 151
RESULT 4
HSTCF1B 2814 bp mRNA PRI 17-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form B).
DEFINITION X59870.X55329
ACCESSION X59870.1 GI:36787
VERSION X59870.1 GI:36787
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2814)
van de Wetering, M.
Direct Submission
Submitted (28-MAR-1991) M. Van De Wetering, Dept. of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
2 (bases 1 to 2814)
van de Wetering, M., Oosterwegel, M., Dooljes, D. and Clevers, H.
IDENTIFICATION AND CLONING OF TCF-1, A T LYMPHOCYTE-SPECIFIC
TRANSCRIPTION FACTOR CONTAINING A SEQUENCE-SPECIFIC HMG BOX
EMBO J. 10 (1), 123-132 (1991)
JOURNAL
MEDLINE 91114695
AUTHORS See also X59869-X59871.
TITLE Location/Qualifiers
FEATURES
SOURCE
1. .2814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat"
/clone_lib="cDNA"
1. .1254
/gene="TCF-1"
/evidence="experimental"
1. .1254
/gene="TCF-1"
80. .889
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1, splice form A"
/protein_id="CAA42526.1"
/db_xref="GI:36786"
/db_xref="SWISS-PROT:P36402"
/translation="MYKTVYSAFNLMLHYPPSGAGHQPQPPPLHKANOPPHGVPO
LSLYEHNSPHPTAPADISQKVHRLQTPDLGFSYLTSGMGLPHTVSWFTHPS
LMLGSGVPGHPAAIPHPATVPSPGKQLPFDRLKTOAESKAKEAKKPTIKKPLNA
FMLYMKEMRAKVIAGCTLKESAINQILGRWHALSREEQAKYVELARKERQLHMLY
PCWSARDNYGKKRRSRKREKHQESTTGNPRELKGNGQESLSMSSSSPA"
misc_feature 539. .769
/gene="TCF-1"
/feature="HMG box"
810
misc_feature
/gene="TCF-1"
/feature="alternative splice site"
BASE COUNT 320 a 391 c 335 g 208 t
ORIGIN
Query Match 100.0%; Score 35; DB 9; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcatgcattaccacccctcgaggcagcaggcag 35
|||||
Db 117 TCATGATTACCCACCCCTCGGAGCAGGGCAG 151
RESULT 5
HSTCF1G 2855 bp DNA PRI 30-SEP-1999
LOCUS Homo sapiens TCF-1 gene.
DEFINITION X63901
ACCESSION X63901.1 GI:36791
VERSION X63901.1 GI:36791
KEYWORDS T-cell transcription factor; transcription factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2855)
van de Wetering, M., Oosterwegel, M., Holsteg, F., Pooyes, D.,
Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.
The human T cell transcription factor-1 gene. Structure,
localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
92235082
2 (bases 1 to 2855)
Van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H.
Extensive alternative splicing and dual promoter usage generate
Tcf-1 protein isoforms with differential transcription control
properties
Mol. Cell. Biol. 16 (3), 745-752 (1996)
96182076
3 (bases 1 to 2855)
Van de Wetering, M.L.
JOURNAL
MEDLINE
AUTHORS
TITLE
```

```

Db      866 TCATGATTACCCACCCTTCGGAGCAGGGCAG 900

RESULT   6
HSTCF1C LOCUS       2910 bp      mRNA          PRI      17-JUN-1991
DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form C).
ACCESSION X59871.X5328
VERSION    X59871.1 GI:36789
KEYWORDS   DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2910)
            van de Wetering,M.
            Direct Submission
            Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
            Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
            The Netherlands
REFERENCE  2 (bases 1 to 2910)
            van de Wetering,M., Oosterwegel,M., Dooljes,D. and Clevers,H.
            Identification and cloning of TCF-1, a T lymphocyte-specific
            transcripion factor containing a sequence-specific HMG box
            EMBO J. 10 (1), 123-132 (1991)
MEDLINE    91114695
COMMENT    See also X59869-X59871.
FEATURES   Location/Qualifiers
            source          1..2910
                        /organism="Homo sapientis"
                        /db_xref="taxon:9606"
                        /cell_type="T lymphocyte"
                        /cell_line="Jurkat and HPB-ALL"
                        /clone_lib="cDNA"
            mRna             1..2910
                        /gene="TCF-1"
                        /evidence=experimental
            gene             1..2910
                        /gene="TCF-1"
                        80..886
            CDS              80..886
                        /codon_start=1
                        /product="T cell factor 1, splice form C"
                        /protein_id="CAA42528.1"
                        /db_xref="GI:36790"
                        /db_xref="SWISS-PROT:P36402"
                        /translation="MYKETYISAFNLLMHYPSPSGAQHPQPPPLIKANOPPHCVGV
                                LSLYEHFNPSHTPADIDSKQVHRPIQTDPDLSGYSLTSGMQLPHTVSWETHWTFNA
                                LMIGSVPGHPAAIPHPAIIVPPSGKELOFPDRNLKTQAEKSAREAKKPITKKPLLA
                                FWLYMKENRAKVIAECTLKESAINQILGRMHALSREEQAKYYELARKEKRLHMQLY
                                PCWSARDKGVKKRRSEKHEQESTTDGPSPKKCRARFGILNQOTDMCGPCR"
            misc_feature      539..769
                        /gene="TCF-1"
                        /note="HMG box"
            misc_feature      810
                        /gene="TCF-1"
                        /note="alternative splice site"
BASE COUNT 782 a 860 c 671 g 597 t
ORIGIN

Query Match           100.0%; Score 35; DB 9; Length 2910;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db	117	TCATGCATTACCCACCCCCCTCGGAGCAGGSCAG	151
RESULT	7		
AC009012/c			
LOCUS	AC009012	84544 bp	DNA
			HTG
			19-APR-2001

Db 38702 TCATGATTACCCACCCCTCGGAGCAGGCAG 38668
|||||
AC011336 86914 bp DNA HTG 23-APR-2001
Homo sapiens chromosome 5 clone CTC-250113, WORKING DRAFT SEQUENCE,
1 ordered pieces.
AC011336
AC011336.4 GI:9256274
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86914)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 86914)
DOE Joint Genome Institute.
Direct Submission
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710593.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 281384, H247
Center clone name: CIT-HSPC_250113

Summary Statistics
Consensus quality: 82614 bases at least Q40
Consensus quality: 85602 bases at least Q30
Consensus quality: 86382 bases at least Q20
Estimated insert size: 89000; pulse field gel estimation
Estimated insert size: 86914; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 6.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86914: contig of 86914 bp in length.
Location/Qualifiers
1. 86914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-250113"
/clone_lib="Caltech human RAC library C"
BASE COUNT 20745 a 22135 c 22612 g 21422 t
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 86914;
Best local Similarity 100.0%; Pred. No. 0.00081;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 45277 TCATGATTACCCACCCCTCGGAGCAGGCAG 45243
|||||
AC011336 86914 bp DNA HTG 23-APR-2001
Homo sapiens chromosome 5 clone CTC-250113, WORKING DRAFT SEQUENCE,
1 ordered pieces.
AC011336
AC011336.4 GI:9256274
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 84544)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 84544)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 14, 2001 this sequence version replaced gi:7454202.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1189075, H71
Center clone name: Xxp1-360D11

Summary Statistics
Consensus quality: 78189 bases at least Q40
Consensus quality: 82001 bases at least Q30
Consensus quality: 83201 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 83944; sum-of-contigs estimation
Quality coverage: 6.34 in Q20 bases; pulse field gel estimation
Quality coverage: 6.04 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2979: contig of 2979 bp in length
* 2980 3079: gap of unknown length
* 3080 34576: contig of 31497 bp in length
* 34577 34676: gap of unknown length
* 34677 59262: contig of 24586 bp in length
* 59263 59362: gap of unknown length
* 59363 76010: contig of 16648 bp in length
* 76011 76110: gap of unknown length
* 76111 79613: contig of 3503 bp in length
* 79614 79713: gap of unknown length
* 79714 80649: contig of 936 bp in length
* 80650 80749: gap of unknown length
* 80750 84544: contig of 3795 bp in length.
Location/Qualifiers
1. 84544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXp1-360D11"
BASE COUNT 19558 a 21740 c 22203 g 20376 t 667 others
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 84544;
Best local Similarity 100.0%; Pred. No. 0.00081;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcatgcattaccacccctcggagcagggcag 35

Center clone name: CIT978SKB_113120

Summary Statistics
Consensus quality: 174733 bases at least Q40
Consensus quality: 166693 bases at least Q30
Consensus quality: 190547 bases at least Q20
Estimated insert size: 165000; pulse field gel estimation
Estimated insert size: 198031; sum-of-contigs estimation
Quality coverage: 7.77 in Q20 bases; pulse field gel estimation
Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.
Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1115: contig of 1115 bp in length
* 1116 1215: gap of unknown length
* 1216 2567: contig of 1352 bp in length
* 2568 2667: gap of unknown length
* 2668 3745: contig of 1078 bp in length
* 3746 3845: gap of unknown length
* 3846 4920: contig of 1075 bp in length
* 4921 5020: gap of unknown length
* 5021 6044: contig of 1024 bp in length
* 6045 6144: gap of unknown length
* 6145 7285: contig of 1141 bp in length
* 7286 7385: gap of unknown length
* 7386 8659: contig of 1274 bp in length
* 8660 8759: gap of unknown length
* 8760 9804: contig of 1045 bp in length
* 9805 11696: contig of 1792 bp in length
* 11697 11799: gap of unknown length
* 11799 12841: contig of 1045 bp in length
* 12842 12941: gap of unknown length
* 12942 15043: contig of 2102 bp in length
* 15044 15143: gap of unknown length
* 15144 17289: contig of 2146 bp in length
* 17290 17389: gap of unknown length
* 17390 20232: contig of 2843 bp in length
* 20233 20332: gap of unknown length
* 20333 21884: contig of 1552 bp in length
* 21885 21984: gap of unknown length
* 21985 24278: contig of 2294 bp in length
* 24279 24378: gap of unknown length
* 24379 28103: contig of 3725 bp in length
* 28104 28203: gap of unknown length
* 28204 30867: contig of 2664 bp in length
* 30868 34914: gap of unknown length
* 34915 35014: contig of 3947 bp in length
* 35015 36030: contig of 1016 bp in length
* 36031 36130: gap of unknown length
* 36131 38852: contig of 2722 bp in length
* 38853 38952: gap of unknown length
* 38953 41871: contig of 2919 bp in length
* 41872 41971: gap of unknown length
* 41972 47080: contig of 5109 bp in length
* 47081 47180: gap of unknown length
* 47181 47550: contig of 2570 bp in length
* 47551 49751: gap of unknown length
* 49752 57118: contig of 7268 bp in length
* 57119 57218: gap of unknown length
* 57219 66046: contig of 8828 bp in length
* 66047 66146: gap of unknown length
* 66147 81046: contig of 14900 bp in length
* 81047 81147: gap of unknown length
* 81147 101650: contig of 20504 bp in length
* 101651 101750: gap of unknown length
* 101751 137108: contig of 35358 bp in length

* 137109 137208: gap of unknown length
* 137209 200831: contig of 63623 bp in length.
FEATURES
source
Location/Qualifiers
1..200831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-113120"
/clone_lib="Caltech human BAC library B"
BASE COUNT 53266 a 46907 c 47140 g 50695 t 2823 others
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 200831;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcatgattaccacccccctcggagcauggcag 35
|||||
Db 184023 TCATGATTACCCACCCCCCTCGGGACGAGCAG 184057
|||||
RESULT 11
MTCSTF MTCSTF 1790 bp mRNA ROD 02-SEP-1991
LOCUS Mouse mRNA for T-cell specific transcription factor.
DEFINITION X61385
ACCESSION X61385.1 GI:54760
VERSION HMG box; T-cell specific transcription factor; transcription
KEYWORDS factor.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1790)
Costerwegel, M.A.
Submitted (12-AUG-1991) M.A. Oosterwegel, Dept of Clin Immunol,
Univ Hospital Ulrecht, P.O. Box 85500, 3508 GA Utrecht, THE
NETHERLANDS
2 (bases 1 to 1790)
Costerwegel, M., van de Wetering, M., Doolijes, D., Klomp, L.,
Winoto, A., Georgopoulos, K., Meijlink, F. and Clevers, H.
Cloning of murine TCF-1, a T-cell-specific transcription factor
interacting with functional motifs in the CD3-epsilon and T cell
receptor alpha enhancers
J. Exp. Med. 173 (5), 1133-1142 (1991)
J. Exp. Med. 173 (5), 1133-1142 (1991)
91217625
Location/Qualifiers
1..1790
/organism="Mus musculus"
/isolate="M2a/M5"
/db_xref="taxon:10090"
/tissue_type="thymus"
/cell_type="T cell"
/clone_lib="lambda Zap cDNA"
1..1790
190..1101
/codon_start=1
/product="T-cell specific transcription factor"
/protein_id="CAA43658.1"
/db_xref="GI:54761"
/db_xref="MGD:MGI:98507"
/db_xref="SWISS-PROT:Q00417"
/translation="MYETVYSAFLNLMIPYPASGAGQHPQPPPLHKKGPQPPHGV
QLSPLYHFSSPHPTAPADISOKGVHRPLQTPDLSGFYSLTSGSMCOIPIHTVSWPS
PPLYPSPSCGYRHFAPTAACAPYRPTHSLMGSGVPHPAIIPHPAIVPSSG
KOELQPYDRNLKTOAFPAKRAKPKVKKPLNAPMLYMKEMKAKVIACTIKESNAI
NOILGRNHIALSREEQAKYVELAKKREQLHMQLYPGWSARDNTGKKRKRREKHQFST
TGGRNAPFTYPEKAAAPAPLPMTVL"
489..581
/note="ACT splice/exon M5"
misc_feature 489..581
misc_feature 761..984

misc_feature 4363..6417
/note="assembly_fragment"
misc_feature 6518..9412
/note="assembly_fragment"
misc_feature 9513..10790
/note="assembly_fragment"
clone_end:SP6
vector_side:right
misc_feature 10891..15080
/note="assembly_fragment"
misc_feature 15181..18824
/note="assembly_fragment"
misc_feature 18925..23129
/note="assembly_fragment"
misc_feature 23230..26859
/note="assembly_fragment"
misc_feature 26960..32575
/note="assembly_fragment"
misc_feature 32676..38781
/note="assembly_fragment"
misc_feature 38882..44764
/note="assembly_fragment"
misc_feature 44865..52115
/note="assembly_fragment"
misc_feature 52216..59927
/note="assembly_fragment"
misc_feature 60028..67115
/note="assembly_fragment"
misc_feature 67216..75188
/note="assembly_fragment"
misc_feature 75289..85641
/note="assembly_fragment"
misc_feature 85742..97849
/note="assembly_fragment"
misc_feature 97950..109572
/note="assembly_fragment"
misc_feature 109673..122441
/note="assembly_fragment"
misc_feature 122542..140243
/note="assembly_fragment"
clone_end:T7
vector_side:left
misc_feature 140344..162496
/note="assembly_fragment"
BASE COUNT 42470 a 37212 c 36404 g 44204 t 2206 others
ORIGIN
Query Match 77.1%; Score 27; DB 2; Length 162496;
Best Local Similarity 85.7%; Pred. No. 0.82;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 tcatgcattaccaccccccctcgaggagcgag 35
||||| ||||||| |||| |||||||||
Db 61998 TCATGCCCTACCCACCGCGCCTCCGGAGAGGGCAG 61964
RESULT 13
AF172532/c 318 bp RNA VRL 29-AUG-1999
LOCUS Hepatitis G virus isolate KZN-B20 5' untranslated region.
DEFINITION
ACCESSION AF172532
VERSION AF172532.1 GI:5802517
KEYWORDS
SOURCE Hepatitis G virus.
ORGANISM Hepatitis G virus.
Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
GBV-C/HGV group.
REFERENCE 1 (bases 1 to 318)
AUTHORS Sathar,M.A., Soni,P.N., Pegoraro,R., Simmonds,P., Smith,D.B.,
Dhillon,A.P. and Dusheiko,G.M.
TITLE A new variant of GB virus C/hepatitis G virus (GBV-C/HGV) from
South Africa
JOURNAL Unpublished
AUTHORS Sathar,M.A., Soni,P.N., Pegoraro,R., Simmonds,P., Smith,D.B.,
Dhillon,A.P. and Dusheiko,G.M.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-1999) Medicine, University of Natal, Faculty of
Medicine, Private Bag 7, Congella, Durban, KwaZulu Natal 4013,
South Africa
FEATURES
Location/Qualifiers
1..318
/organism="Hepatitis G virus"
/isolate="KZN-B20"
/db_xref="taxon:45255"
/country="South Africa"
1..318
/note="polyprotein"
BASE COUNT 61 a 99 c 105 g 53 t
ORIGIN
5'UTR
Query Match 60.6%; Score 21.2; DB 14; Length 318;
Best Local Similarity 88.5%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 tcatgcattaccaccccccctcgaggga 26
||||||| ||| |||||| |||||
Db 309 TCATGCATTTCCCGCCCCCGGGGA 284

misc_feature 4363..6417
/note="assembly_fragment"
misc_feature 6518..9412
/note="assembly_fragment"
misc_feature 9513..10790
/note="assembly_fragment"
clone_end:SP6
vector_side:right
misc_feature 10891..15080
/note="assembly_fragment"
misc_feature 15181..18824
/note="assembly_fragment"
misc_feature 18925..23129
/note="assembly_fragment"
misc_feature 23230..26859
/note="assembly_fragment"
misc_feature 26960..32575
/note="assembly_fragment"
misc_feature 32676..38781
/note="assembly_fragment"
misc_feature 38882..44764
/note="assembly_fragment"
misc_feature 44865..52115
/note="assembly_fragment"
misc_feature 52216..59927
/note="assembly_fragment"
misc_feature 60028..67115
/note="assembly_fragment"
misc_feature 67216..75188
/note="assembly_fragment"
misc_feature 75289..85641
/note="assembly_fragment"
misc_feature 85742..97849
/note="assembly_fragment"
misc_feature 97950..109572
/note="assembly_fragment"
misc_feature 109673..122441
/note="assembly_fragment"
misc_feature 122542..140243
/note="assembly_fragment"
clone_end:T7
vector_side:left
misc_feature 140344..162496
/note="assembly_fragment"
BASE COUNT 42470 a 37212 c 36404 g 44204 t 2206 others
ORIGIN
Query Match 77.1%; Score 27; DB 2; Length 162496;
Best Local Similarity 85.7%; Pred. No. 0.82;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 tcatgcattaccaccccccctcgaggagcgag 35
||||| ||||||| |||| |||||||||
Db 61998 TCATGCCCTACCCACCGCGCCTCCGGAGAGGGCAG 61964
RESULT 13
AF172532/c 318 bp RNA VRL 29-AUG-1999
LOCUS Hepatitis G virus isolate KZN-B20 5' untranslated region.
DEFINITION
ACCESSION AF172532
VERSION AF172532.1 GI:5802517
KEYWORDS
SOURCE Hepatitis G virus.
ORGANISM Hepatitis G virus.
Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
GBV-C/HGV group.
REFERENCE 1 (bases 1 to 318)
AUTHORS Sathar,M.A., Soni,P.N., Pegoraro,R., Simmonds,P., Smith,D.B.,
Dhillon,A.P. and Dusheiko,G.M.
TITLE A new variant of GB virus C/hepatitis G virus (GBV-C/HGV) from
South Africa
JOURNAL Unpublished
AUTHORS Sathar,M.A., Soni,P.N., Pegoraro,R., Simmonds,P., Smith,D.B.,
Dhillon,A.P. and Dusheiko,G.M.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-1999) Medicine, University of Natal, Faculty of
Medicine, Private Bag 7, Congella, Durban, KwaZulu Natal 4013,
South Africa
FEATURES
Location/Qualifiers
1..318
/organism="Hepatitis G virus"
/isolate="KZN-B20"
/db_xref="taxon:45255"
/country="South Africa"
1..318
/note="polyprotein"
BASE COUNT 61 a 99 c 105 g 53 t
ORIGIN
5'UTR
Query Match 60.6%; Score 21.2; DB 14; Length 318;
Best Local Similarity 88.5%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 tcatgcattaccaccccccctcgaggga 26
||||||| ||| |||||| |||||
Db 309 TCATGCATTTCCCGCCCCCGGGGA 284

Thu Feb 21 07:25:44 2002

us-09-904-420a-1_copy_866_900.rge

```
RESULT 15
HSVDAC2      2321 bp      DNA      PRI      19-OCT-1999
LOCUS        Homo sapiens voltage-dependent anion channel (VDAC1) gene, exons 2
DEFINITION   and 3.
ACCESSION    AF151093
VERSION      AF151093.1 GI:5880853
KEYWORDS
SEGMENT      2 of 6
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 2321)
              Decker,W.K., Bowles,K.R., Schatte,E.C., Towbin,J.A. and
              Craigen,W.J.
              Revised fine mapping of the human voltage-dependent anion channel
              loci by radiation hybrid analysis
              Mamm. Genome 10 (10), 1041-1042 (1999)
JOURNAL      99431679
MEDLINE      10501981
PUBMED
REFERENCE    2 (bases 1 to 2321)
AUTHORS      Decker,W.K., Bowles,K.R., Schatte,E.C., Towbin,J.A. and
              Craigen,W.J.
              Direct Submission
JOURNAL      Submitted (14-MAY-1999) Molecular and Human Genetics, Baylor
              College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
FEATURES
source       1..2321
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="5"
              /map="5q32"
              1209..1275
              /gene="VDAC1"
              /number=2
              1866..1915
              /gene="VDAC1"
              /number=3
BASE COUNT   582 a 452 c 556 g 731 t
ORIGIN
```

Query Match 60.6%; Score 21.2; DB 9; Length 2321;
Best Local Similarity 76.5%; Pred. No. 3.7e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccccctcggagcaggca 34
||||| ||| | ||| | ||| |||
Db 259 TCATGCATGACCCCTCTCTCTGCGGACAGCA 292

Search completed: February 20, 2002, 14:22:05
Job time: 14164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 11:03:31 ; Search time 366.07 Seconds
(without alignments)
81.969 Million cell updates/sec

Title: US-09-904-420a-1_COPY_866_900
Perfect score: 35
Sequence: 1 tcatgcattaccacccctcggagcgggagcag 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
C 1	19.8	56.6	299	20 AAV00834
C 2	19.8	56.6	3500	21 AAV00812
C 3	19.8	56.6	3803	22 AAV00788
C 4	19.8	56.6	3804	22 AAV00788
C 5	19.6	56.0	512	19 AAV0002
C 6	19.6	56.0	1565	13 AAV04840
C 7	19.6	56.0	8210	20 AAV031725
C 8	19.6	56.0	8239	18 AAV04534
C 9	19.2	54.9	1704	21 AAV02350
C 10	19	54.3	41	14 AAV03563
C 11	19	54.3	41	17 AAV07727

12	19	54.3	41	18 AAV00834
13	19	54.3	41	18 AAV00812
14	19	54.3	41	19 AAV14619
15	19	54.3	41	19 AAV14597
16	19	54.3	41	20 AAV79694
17	19	54.3	41	20 AAV79672
18	19	54.3	41	21 AAV92772
19	19	54.3	41	21 AAV92794
20	19	54.3	2000	18 AAT58391
21	19	54.3	2000	20 AAV84063
C 22	19	54.3	3861	20 AAX60261
C 23	19	54.3	66986	22 AAF28542
C 24	18.8	53.7	7661	22 AAF28542
25	18.8	53.7	7661	22 AAF28542
26	18.8	53.7	7661	22 AAF28542
27	18.8	53.7	7661	22 AAF28542
28	18.8	53.7	7661	22 AAF28542
29	18.8	53.7	7661	22 AAF28542
30	18.8	53.7	7661	22 AAF28542
C 31	18.6	53.1	461	22 AAF28542
C 32	18.6	53.1	461	22 AAF28542
C 33	18.6	53.1	461	22 AAF28542
C 34	18.6	53.1	461	22 AAF28542
C 35	18.6	53.1	461	22 AAF28542
C 36	18.6	53.1	461	22 AAF28542
C 37	18.6	53.1	461	22 AAF28542
C 38	18.4	52.6	1611	21 AAC77571
C 39	18.4	52.6	1611	21 AAC77571
C 40	18.4	52.6	1611	21 AAC77571
C 41	18.4	52.6	1611	21 AAC77571
C 42	18.4	52.6	1611	21 AAC77571
C 43	18.4	52.6	1611	21 AAC77571
C 44	18.4	52.6	1611	21 AAC77571
C 45	18.4	52.6	1611	21 AAC77571

ALIGNMENTS

RESULT 1
AAZ13639/c
AAZ13639 standard; cDNA; 299 BP.
XX ID AAZ13639
XX AC AAZ13639
XX DT 12-OCT-1999 (first entry)
XX DE Human gene expression product cDNA sequence SEQ ID NO:1108.
XX KW Human; gene; gene expression product; diagnosis; therapy; probe;
XX KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX OS Homo sapiens.
XX PN WO9338972-A2.
XX PD 05-AUG-1999.
XX PF 28-JAN-1999; 99WO-US01619.
XX PR 03-APR-1998; 98US-0080666.
XX PR 28-JAN-1998; 98US-0072910.
XX PR 24-FEB-1998; 98US-0075954.
XX PR 31-MAR-1998; 98US-0080114.
XX PR 03-APR-1998; 98US-0080515.
XX PA (CHIR) CHIRON CORP.
XX PA (HYSE-) HYSEQ INC.
XX PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

05-OCT-2000.
31-MAR-2000: 2000WO-US08621.
31-MAR-1999: 99US-0127607.
02-APR-1999: 99US-0127636.
05-APR-1999: 99US-0127728.
30-MAR-2000: 2000US-0540763.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M:
WPI: 2000-602362/57.
P-PSDB: AAB40726.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 5; Page 957-959; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnary;
antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
osteoplastic; anticonvulsant; antiarthritic; coagulant; immunosuppressant;
immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antitumor; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.
Sequence 3500 BP; 604 A; 1155 C; 1093 G; 646 T; 2 other:
Query Match 56.6%; Score 19.8; DB 21; Length 3500;
Best Local Similarity 77.4%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 3 atgcattaccacccctcgaggagcaggc 33
||||| ||||| ||||| ||||| |||||
Db 1446 atgctgtgccactccctcgaggagcaggc 1476
RESULT 3
AA160788/c
ID AA160788 standard; cDNA: 3803 BP.
XX AA160788;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4777.
DE
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemolatic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukaemia; ss.
XX
XX

05-OCT-2000.
31-MAR-2000: 2000WO-US08621.
31-MAR-1999: 99US-0127607.
02-APR-1999: 99US-0127636.
05-APR-1999: 99US-0127728.
30-MAR-2000: 2000US-0540763.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M:
WPI: 2000-602362/57.
P-PSDB: AAB40726.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 5; Page 957-959; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnary;
antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
osteoplastic; anticonvulsant; antiarthritic; coagulant; immunosuppressant;
immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antitumor; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.
Sequence 3500 BP; 604 A; 1155 C; 1093 G; 646 T; 2 other:
Query Match 56.6%; Score 19.8; DB 21; Length 3500;
Best Local Similarity 77.4%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 3 atgcattaccacccctcgaggagcaggc 33
||||| ||||| ||||| ||||| |||||
Db 1446 atgctgtgccactccctcgaggagcaggc 1476
RESULT 3
AA160788/c
ID AA160788 standard; cDNA: 3803 BP.
XX AA160788;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4777.
DE
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemolatic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukaemia; ss.
XX
XX

Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
Stache-Crain B, Sudduth-Klinger J, Williams LT;
WPI: 1999-494092/41.
Novel human genes and their expression products which are
differentially expressed in different cell types
Claim 1; Page 914; 2479pp; English.
The present invention describes a library of human polynucleotides
comprising the sequences given in AA212532 to AA217779. Also described is
a method of detecting differentially expressed genes correlated with the
cancerous state of a mammalian cell, comprising detecting at least one
differentially expressed gene product in a test sample from a cell
suspected of being cancerous, where the gene product is encoded by one
of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
polynucleotides can be used as a source of primers and probes, which can
be used for a variety of purpose, e.g. detection of expression levels,
mapping, tissue typing or profiling, forensics, genetic analysis and
detection of polymorphisms. Polypeptides encoded by the polynucleotides
can be used for raising antibodies for experimental, diagnostic and
therapeutic purposes. The polynucleotides may also be used to construct
arrays for diagnostics (which may be used to determine function of an
encoded protein); and to detect differences in expression levels between
two cells (e.g. to identify abnormal or diseased tissue in a human, to
identify a genetic predisposition or susceptibility to a disease such as
cancer). The polynucleotides of the invention are especially used in the
diagnosis, prognosis and management of colorectal cancer, breast cancer,
and lung cancer. The polynucleotides can also be used to screen for
peptide analogues and antagonists.
Sequence 299 BP; 55 A; 99 C; 68 G; 62 T; 15 other:
Query Match 56.6%; Score 19.8; DB 20; Length 299;
Best Local Similarity 77.4%; Pred. No. 58;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 3 atgcattaccacccctcgaggagcaggc 33
||||| ||||| ||||| ||||| |||||
Db 127 ACGCACTACCTCCCTCCCTCGAGGAGCGGC 97
RESULT 2
AAC74935
ID AAC74935 standard; cDNA: 3500 BP.
XX AAC74935;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF490 polynucleotide sequence SEQ ID NO: 979.
XX Human; open reading frame; ORFX: detection; cytostatic; hepatotropic;
KW vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteoplastic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antitumor;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
OS Homo sapiens.
PN WO200058473-A2.
XX


```
XX KW Fc receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
XX KW immune complex related disease; systemic lupus erythematosus; allergy;
XX KW haemolytic anaemia; thrombocytopaenia; anaphylaxis; cancer; lymphoma;
XX KW leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
XX KW FcR-IV; ss.
XX OS Homo sapiens.
XX PN WO9831806-A2.
XX PD 23-JUL-1998.
XX PF 20-JAN-1998; 98WO-US01184.
XX PR 18-JUN-1997; 97US-0049872.
XX PR 21-JAN-1997; 97US-0034205.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Gentz RL, Murphy M, Ni J, Olsen HS, Ruben SM;
XX WPI: 1998-414105/35.
XX PT Nucleic acid encoding Fc receptor-like polypeptides or their
XX PT fragments - and related vectors, transformed cells and antibodies,
XX PT useful for treating and diagnosing diseases of the haematopoietic
XX PT and immune systems
XX PS Claim 38; Page 96-97; 141pp; English.
XX CC This sequence represents a cDNA clone related to the DNA encoding the
XX CC Fc receptor-like IV protein (FcR-IV) of the invention. Cells containing
XX CC the DNA are used to express the recombinant protein, and to screen for
XX CC specific (ant)agonists. The proteins are used to induce phagocytosis, and
XX CC their (ant)agonists are used to treat immune complex related diseases
XX CC (e.g. rheumatoid arthritis, systemic lupus erythematosus, haemolytic
XX CC anaemia, thrombocytopaenia, anaphylaxis, allergy, colorectal or breast
XX CC cancer, lymphoma, leukaemia, infection by intracellular pathogens etc).
XX CC The antagonists are also useful as immunomodulators and inhibitors of
XX CC viral (e.g. human immune deficiency or dengue viruses) entry into cells.
XX CC The proteins may also be used to screen for specific binding agents.
XX CC i.e. (ant)agonists, for raising antibodies (Ab), and for identification
XX CC of particular cells or tissues. The Ab can be used therapeutically as
XX CC antagonists; as assay reagents for diagnostic determination of the levels
XX CC of expression of the proteins and for affinity purification of the
XX CC proteins. The DNA and its fragments are useful as hybridisation probes or
XX CC primers for isolating related genes, in situ hybridisation (chromosome
XX CC mapping) and diagnostically to measure mRNA expression.
XX SQ Sequence 512 BP; 111 A; 121 C; 153 G; 102 T; 25 other;

Query Match 56.0%; Score 19.6; DB 19; Length 512;
Best Local Similarity 73.5%; Pred. No. 73;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 catgcattaccaccccccctggagcaggcag 35
| | | | | | | | | | | | | | | | | | | |
Db 321 cctgcattcgtggcccccttggagcaggcag 354

RESULT 6
AAQ31725
ID AAQ31725 standard; DNA; 1565 BP.
XX AC AAQ31725;
XX DT 04-APR-1993 (first entry)
XX DE Mouse Igh 3'-enhancer.
XX KW Rat; cross hybridisation; B cell specific; target; ss.

Query Match 56.0%; Score 19.6; DB 19; Length 512;
Best Local Similarity 73.5%; Pred. No. 73;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 catgcattaccaccccccctggagcaggcag 35
| | | | | | | | | | | | | | | | | | | |
Db 321 cctgcattcgtggcccccttggagcaggcag 354

RESULT 6
AAQ31725
ID AAQ31725 standard; DNA; 1565 BP.
XX AC AAQ31725;
XX DT 04-APR-1993 (first entry)
XX DE Mouse Igh 3'-enhancer.
XX KW Rat; cross hybridisation; B cell specific; target; ss.
```

```
XX OS Mus musculus.
XX PN WO9221762-A.
XX PD 10-DEC-1992.
XX PF 03-JUN-1992; 92WO-SE00375.
XX PR 07-JUN-1991; 91SE-0001740.
XX PA (PETT/) PETTERSSON S.
XX PI Pettersson S;
XX WPI: 1992-433660/52.
XX PT Mouse immunoglobulin H 3' enhancer - which cross-hybridises with
XX PT rat 3' enhancer and is used to target tissue-specific gene
XX PT expression
XX PS Claim 2; Fig 3A; 33pp; English.
XX CC A mouse liver library in bacteriophage was screened using as a probe
XX CC an AccI-BglI subfragment from the core of the rat 3' enhancer
XX CC (nucleotides 432-748). Two overlapping gps. of clones (lambda M2 and
XX CC lambda M3) were isolated. Phage lambda M3 extends 3' of lambda M2.
XX CC The sequence of the mouse Igh 3' enhancer was determined and aligned
XX CC with that of the rat. Hybridisations of subclones of the mouse Igh
XX CC 3' enhancer to phage lambda M2 showed that the 3' enhancers of mouse
XX CC and rat were present in the genome in opposite orientations. The
XX CC enhancers are B cell specific and may be used to target tissue
XX CC specific expression of genes, so may be useful therapeutically, e.g.
XX CC in targeting prodn. of proteins and in hybridoma technology. The
XX CC enhancer may be used to enhance expression of genes in host cells, in
XX CC vivo or in vitro, partic. in certain lymphoid cell lines and in
XX CC transgenic animals. It may be used for the prodn. of monoclonal
XX CC antibodies. See also AAQ31726.
XX SQ Sequence 1565 BP; 405 A; 388 C; 460 G; 312 T; 0 other;

Query Match 56.0%; Score 19.6; DB 13; Length 1565;
Best Local Similarity 73.5%; Pred. No. 78;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 catgcattaccaccccccctggagcaggcag 35
| | | | | | | | | | | | | | | | | | | |
Db 1192 catccattaccacccaccactggatggatggcag 1225

RESULT 7
AAQ04534
ID AAQ04534 standard; DNA; 8210 BP.
XX AC AAQ04534;
XX DT 13-APR-1999 (first entry)
XX DE DNA encoding human ATR protein.
XX KW Alaxia telangiectasia; ATM protein; assay; interaction; kinase activity;
XX KW p53; screening; ATR; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 106..8040
XX FT /*tag= a
XX FT /product= ATR
XX PN GB2327498-A.
```

PD	27-JAN-1999.
XX	
XX	16-JUL-1998; 98GB-0015423.
PF	
XX	
PR	16-JUL-1997; 97GB-0014971.
XX	(CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
XX	Jackson SP, Lakin ND, Smith GCM;
PI	WPI; 1999-073587/07.
XX	P-PSDB; AAW84271.
DR	
XX	Assay method for compounds modulating the interaction of ATM and p53
PT	- useful for the treatment of e.g. cancer, immunosuppression and HIV
PT	infections and for the purification of the proteins ATM and ATR
XX	Disclosure; Fig 8b; 124pp; English.
XX	The present sequence encodes a human ATR (Frp1) protein. The protein is
CC	used in the course of the invention. The specification describes an
CC	assay method for a compound able to modulate the interaction between
CC	ATM or a protein having an associated kinase activity and p53 or a
CC	protein having homologous phosphorylation sites. The assay comprises
CC	contacting a peptide fragment ATM with a relevant fragment of p53
CC	and a test compound, and determining the interaction or binding
CC	between the substances and the test compound. The assay method is
CC	useful for screening for compounds able to modulate the interaction
CC	between ATM and p53. The screened agents, peptide fragments and
CC	nucleic acids are useful for therapy involving modulating ATM action
CC	e.g. in the treatment of cancer, immunosuppression or HIV infections by
CC	modulating phosphorylation of p53 by ATM, and for purifying the proteins
CC	ATM and ATR.
XX	
SQ	Sequence 8210 BP; 2511 A; 1555 C; 1738 G; 2406 T; 0 other;
Query Match 56.0%; Score 19.6; DB 20; Length 8210;	
Best Local Similarity 73.5%; Pred No. 87;	
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
OY	2 catgcattaccacccccctcggagcaggcgag 35
Db	131 ctccatgatcccgcctcgggagctggggcgag 164
RESULT 8	
AAT61981	
ID	AAT61981 standard; cDNA; 8239 BP.
AC	
XX	AAT61981;
DT	
XX	01-JUL-1997 (first entry)
DE	Human ataxia and rad related checkpoint protein cdna.
XX	
KW	Ataxia and rad related checkpoint protein; ATR gene; cell cycle;
KW	Cancer; therapy; ss.
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 80..8014
FT	/tag= a
XX	
PN	WC9709433-Al.
XX	
PD	13-MAR-1997.
XX	
PP	06-SEP-1996; 96WO-GB02197.
XX	
PR	06-SEP-1995; 95GB-0018220.
XX	

Key	Location/Qualifiers
stem_loop	12..26
misc_binding	/*tag= a
misc_binding	12..16
misc_binding	/*tag= b
misc_binding	/*note= "stem 1(a) binds to stem 1(b) (bases 22-26)"
misc_binding	22..26
stem_loop	/*tag= c
stem_loop	/*note= "stem 1(b) binds to stem 1(a) (bases 12-16)"
stem_loop	17..34
misc_binding	/*tag= d
misc_binding	17..21
misc_binding	/*tag= e
misc_binding	/*note= "stem 2(a) binds to stem 2(b) (bases 30-34)"
misc_binding	30..34
misc_binding	/*tag= f
misc_binding	/*note= "stem 2(b) binds to stem 2(a) (bases 17-21)"
US5475096-A.	
12-DEC-1995.	
11-JUN-1990;	90US-0536428.
10-JUN-1991;	91US-0714131.
11-JUN-1990;	90US-0536428.
(OYRE-) UNIV RES CORP.	
Gold L, Tuerk C;	
WPI; 1996-039557/04.	
Artificial nucleic acid ligands - for selected target proteins.	
Example 2; Column 171-172; 133pp; English.	
AA07716-T07740 represent HIV-1 reverse transcriptase (HIV-1 RT) ligands identified by a systematic evolution of ligands by exponential enrichment (SELEX) reaction. This sequence forms a pseudoknot structure (a double stem loop). The starting sequence for this SELEX reaction is represented by AA07715. In a SELEX reaction, a target molecule is contacted with a mixture of random nucleic acids under conditions favourable for binding. Unbound nucleic acids are then separated from those bound to the target, and the nucleic acid-target pairs are dissociated. The dissociated nucleic acids are amplified to give a ligand enriched mixture. These steps are repeated until the specific ligand is obtained. This procedure can also be carried out for ligands for bacteriophage coat proteins, serine proteases, mammalian receptors, mammalian hormones, mammalian growth factors, ribosomal proteins, DNA polymerases and viral rev proteins. The ligands identified (such as this sequence) may be used in assays, diagnostic procedures, or cell sorting as an inhibitor of the target molecule function. It may also be used as a probe or sequestering agent, and also possess catalytic activity.	
Sequence 41 BP; 13 A; 9 C; 13 G; 6 U; 0 other;	
Query Match	54.3%; Score 19; DB 17; Length 41;
Best Local Similarity	62.9%; Pred. No. 1.1e+02;
Matches	22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
1 tcatgcattaccacccctccgcgagcaggcgag 35	
: : : : : :	
1 ucaaggaauaacccgacgcacaggaggaugcgag 35	
RESULT 12	
AAV00834	
ID	AAV00834 standard; RNA; 41 BP.
XX	
AAV00834:	

Thu Feb 21 07:25:44 2002

```
XX PD 23-SEP-1997.
XX PF 27-MAR-1995; 95US-0412110.
XX PR 10-JUN-1991; 91US-0714131.
XX PR 11-JUN-1990; 90US-0536428.
XX PR 27-MAR-1995; 95US-0412110.
XX PA (NEXS-) NEXSTAR PHARM INC.
XX PI Gold L, Tuerk C;
XX DR WPI; 1997-479527/44.
XX PT Nucleic acid ligands for binding proteins - obtained by systematic
XX PT evolution of ligands by exponential enrichment procedures
XX PS Example 2; Column 55-56; 133pp; English.
XX CC This sequence represents a nucleic acid ligand which binds the human
XX CC immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT).
XX CC Ligands to the HIV-1 RT are isolated by the systematic evolution of
XX CC ligands by exponential enrichment (SELEX) method of the invention.
XX CC This method is especially used to isolate novel non-naturally occurring
XX CC nucleic acid ligands having a specific binding affinity for a target
XX CC molecule, where the target molecule is a protein and the nucleic acid
XX CC ligand is not a nucleic acid known to bind the target molecule. The
XX CC nucleic acid ligands can be used, e.g. in assay methods, diagnostic
XX CC procedures, cell sorting, as inhibitors of target molecule function,
XX CC as probes, as sequestering agents, for therapy or as catalysts.
XX SQ Sequence 41 BP; 13 A; 9 C; 13 G; 6 U; 0 other;

Query Match 54.3%; Score 19; DB 18; Length 41;
Best Local Similarity 62.9%; Pred. No. 1.1e+02;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcattgattaccacccctcggagcaggcgca 35
Db 1 ucaaggauuaccgacgcacgaaggcgca 35

RESULT 14
AAV14619
ID AAV14619 standard; RNA; 41 BP.
XX AC AAV14619;
XX DT 21-MAY-1998 (first entry)
XX DE Pseudoknot 1.8 of SELEX identified ligand for DNA polymerase.
XX KW High affinity RNA ligand motif; polymer binding; cell sorting; inhibitor;
XX KW systematic evolution of ligands by exponential enrichment; SELEX;
XX KW sequestering agent; DNA polymerase; ss.
XX OS Synthetic.
XX OS Bacteriophage t4.
XX PN US5696249-A.
XX PD 09-DEC-1997.
XX PF 24-MAR-1995; 95US-0409442.
XX PR 10-JUN-1991; 91US-0714131.
XX PR 11-JUN-1990; 90US-0536428.
XX PR 24-MAR-1995; 95US-0409442.
XX PA (NEXS-) NEXSTAR PHARM INC.
XX PI Gold L, Tuerk C;
XX DR WPI; 1998-041356/04.

Gold L, Tuerk C;
WPI; 1998-041356/04.
Synthetic nucleic acid ligands - that bind to target molecules other
than nucleic acids
Example 1; Column 55-56; 137pp; English.
This sequence represents the pseudoknot portion of the random region of a
ligand identified by a systematic evolution of ligands by exponential
enrichment (SELEX) reaction to isolate ligands specific for bacteriophage
T4 DNA polymerase. This sequence is part of a ligand of the invention.
The ligands are non-naturally occurring nucleic acid
ligand with specific binding affinity for a target molecule, where: the
target molecule is not a polynucleotide that binds to the ligand by
Watson-Crick base pairing or triple helix binding; the ligand is not a
nucleic acid having the known physiological function of being bound by
the target molecule; and the ligand is obtained by: (a) contacting the
target molecule with a candidate mixture of nucleic acids, each having a
region of randomised sequence; (b) separating the nucleic acids having
the highest affinity for the target; and (c) amplifying the separated
nucleic acids. ligands as above that bind to natural or synthetic
polymers, e.g. proteins, polysaccharides, glycoproteins, hormones,
receptors, cell surfaces, metabolites, cofactors, transition-state
analogues or toxins, may be useful in assays, diagnostic procedures or
cell sorting, as inhibitors of target molecule function, as probes, as
sequestering agents, etc., or may have catalytic activity.
Sequence 41 BP; 13 A; 9 C; 13 G; 6 U; 0 other;

Query Match 54.3%; Score 19; DB 19; Length 41;
Best Local Similarity 62.9%; Pred. No. 1.1e+02;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcattgattaccacccctcggagcaggcgca 35
Db 1 ucaaggauuaccgacgcacgaaggcgca 35

RESULT 15
AAV14597
ID AAV14597 standard; RNA; 41 BP.
XX AC AAV14597;
XX DT 21-MAY-1998 (first entry)
XX DE Random region 1.8 of SELEX identified ligand for DNA polymerase.
XX KW High affinity RNA ligand motif; polymer binding; cell sorting; inhibitor;
XX KW systematic evolution of ligands by exponential enrichment; SELEX;
XX KW sequestering agent; DNA polymerase; ss.
XX OS Synthetic.
XX OS Bacteriophage t4.
XX PN US5696249-A.
XX PD 09-DEC-1997.
XX PF 24-MAR-1995; 95US-0409442.
XX PR 10-JUN-1991; 91US-0714131.
XX PR 11-JUN-1990; 90US-0536428.
XX PR 24-MAR-1995; 95US-0409442.
XX PA (NEXS-) NEXSTAR PHARM INC.
XX PI Gold L, Tuerk C;
XX DR WPI; 1998-041356/04.
```


us-09-904-420a-1_copy_866_900.rni

Thu Feb 21 07:25:45 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 10:32:21 ; Search time 162.84 Seconds
(without alignments)
48.678 Million cell updates/sec

Title: US-09-904-420a-1_COPY_866_900
Perfect score: 35
Sequence: 1 catgattaccacccccctcgaggagcaggcag 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6C_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	54.9	8520	4	US-09-272-032-7
2	19	54.3	41	1	Sequence 7, Appli
3	19	54.3	41	1	Sequence 72, Appl
4	19	54.3	41	1	Sequence 94, Appl
5	19	54.3	41	1	Sequence 72, Appl
6	19	54.3	41	1	Sequence 94, Appl
7	19	54.3	41	1	Sequence 72, Appl
8	19	54.3	41	1	Sequence 94, Appl
9	19	54.3	41	1	Sequence 72, Appl
10	19	54.3	41	2	Sequence 94, Appl
11	19	54.3	41	2	Sequence 72, Appl
12	19	54.3	41	3	Sequence 94, Appl
13	19	54.3	41	3	Sequence 72, Appl
C 14	18.6	53.1	4403765	4	US-09-103-840A-2
C 15	18.6	53.1	4411529	4	US-09-103-840A-1
C 16	18.4	52.6	4002	2	US-08-231-193A-53
C 17	18.4	52.6	4002	2	US-08-486-273A-53
C 18	18.4	52.6	4002	3	US-08-480-474-53
C 19	18.4	52.6	4002	3	US-08-940-086A-53
C 20	18.4	52.6	4017	2	US-08-231-193A-49
C 21	18.4	52.6	4017	2	US-08-486-273A-49
C 22	18.4	52.6	4017	3	US-08-480-474-49
C 23	18.4	52.6	4017	3	US-08-940-086A-49
C 24	18.4	52.6	4053	2	US-08-231-193A-47
C 25	18.4	52.6	4053	2	US-08-486-273A-47
C 26	18.4	52.6	4053	3	US-08-480-474-47
C 27	18.4	52.6	4053	3	US-08-940-086A-47

C 28	18.4	52.6	4068	2	US-08-231-193A-5	Sequence 5, Appli
C 29	18.4	52.6	4068	2	US-08-486-273A-5	Sequence 5, Appli
C 30	18.4	52.6	4068	3	US-08-480-474-5	Sequence 5, Appli
C 31	18.4	52.6	4068	3	US-08-940-086A-5	Sequence 5, Appli
C 32	18.4	52.6	4077	2	US-08-231-193A-51	Sequence 51, Appl
C 33	18.4	52.6	4077	2	US-08-486-273A-51	Sequence 51, Appl
C 34	18.4	52.6	4077	3	US-08-480-474-51	Sequence 51, Appl
C 35	18.4	52.6	4077	3	US-08-940-086A-51	Sequence 51, Appl
C 36	18.4	52.6	4092	2	US-08-231-193A-45	Sequence 45, Appl
C 37	18.4	52.6	4092	2	US-08-486-273A-45	Sequence 45, Appl
C 38	18.4	52.6	4092	3	US-08-480-474-45	Sequence 45, Appl
C 39	18.4	52.6	4092	3	US-08-940-086A-45	Sequence 45, Appl
C 40	18	51.4	401	1	US-08-466-033-213	Sequence 213, App
C 41	18	51.4	401	1	US-08-466-033-214	Sequence 214, App
C 42	18	51.4	401	1	US-08-466-033-215	Sequence 215, App
C 43	18	51.4	401	1	US-08-444-733-213	Sequence 213, App
C 44	18	51.4	401	1	US-08-444-733-214	Sequence 214, App
C 45	18	51.4	401	1	US-08-444-733-215	Sequence 215, App

ALIGNMENTS

RESULT 1
US-09-272-032-7/c
; Sequence 7, Application US/09272032A
; Patent No. 6296852
; GENERAL INFORMATION:
; APPLICANT: Johnson, Michael A.
; APPLICANT: McCoy, Richard J.
; TITLE OF INVENTION: Recombinant Avian Adenovirus Vector
; FILE REFERENCE: 48-95a
; CURRENT APPLICATION NUMBER: US/09/272,032A
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 08/448,617
; EARLIER FILING DATE: 1995-09-08
; EARLIER APPLICATION NUMBER: PCT/AU94/000189
; EARLIER FILING DATE: 1994-04-14
; EARLIER APPLICATION NUMBER: AU PL 8297
; EARLIER FILING DATE: 1993-04-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 8520
; TYPB: DNA
; ORGANISM: fowl adenovirus
; US-09-272-032-7

Query Match 54.9%; Score 19.2; DB 4; Length 8520;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 2 catgattaccacccccctcgaggagcaggc 33
||||| | | | | | | | | | | | | | | | | |
Db 3479 CATCCATTTCCTCCGCCCTCTCGGACGAGTC 3448

RESULT 2
US-07-931-473B-72
; Sequence 72, Application US/07931473B
; Patent No. 5270163
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 335
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bealon & Swanson, P.C.
; STREET: 4582 South Ulster Street Parkway, #403
; CITY: Denver
; STATE: Colorado

Thu Feb 21 07:25:45 2002

TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-931-473B-94

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccccctcgaggagcagcag 35
:||| |::| | | | | | | | | | | | | | |
Db 1 UCAAGGAUUAACCGACGCGCAACGGGAGAAUGGCAG 35

RESULT 4
US-07-714-131C-72
; Sequence 72, Application US/07714131C
; Patent No. 5475096
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 344
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beaton & Swanson, P.C.
; STREET: 4582 South Ulster Street Parkway, #403
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80237
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/931.473B
; FILING DATE: 19920817
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 850-9900
; TELEFAX: (303) 850-9401
; INFORMATION FOR SEQ ID NO: 72:
; LENGTH: 41 nucleotides
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-931-473B-72

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccccctcgaggagcagcag 35
:||| |::| | | | | | | | | | | | | | |
Db 1 UCAAGGAUUAACCGACGCGCAACGGGAGAAUGGCAG 35

RESULT 3
US-07-931-473B-94
; Sequence 94, Application US/07931473B
; Patent No. 5270163
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 335
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beaton & Swanson, P.C.
; STREET: 4582 South Ulster Street Parkway, #403
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80237
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/931.473B
; FILING DATE: 19920817
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 850-9900
; TELEFAX: (303) 850-9401
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides

TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-931-473B-94

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccccctcgaggagcagcag 35
:||| |::| | | | | | | | | | | | | | |
Db 1 UCAAGGAUUAACCGACGCGCAACGGGAGAAUGGCAG 35

RESULT 4
US-07-714-131C-72
; Sequence 72, Application US/07714131C
; Patent No. 5475096
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 344
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beaton & Swanson, P.C.
; STREET: 4582 South Ulster Street Parkway, #403
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80237
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/714.131C
; FILING DATE: June 10, 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 850-9900
; TELEFAX: (303) 850-9401
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-714-131C-72

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccccctcgaggagcagcag 35
:||| |::| | | | | | | | | | | | | | |
Db 1 UCAAGGAUUAACCGACGCGCAACGGGAGAAUGGCAG 35

RESULT 5
US-07-714-131C-94
; Sequence 94, Application US/07714131C
; Patent No. 5475096
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk

; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 344
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beaton & Swanson, P.C.
; STREET: 4582 South Ulster Street Parkway, #403
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80237
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/714.131C
; FILING DATE: June 10, 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 850-9900
; TELEFAX: (303) 850-9401
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-714-131C-94

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Oy 1 tcatgcattaccacccccctcgaggagcgagcag 35
:|||||:|||||:|||||:|||||
Db 1 UCAAGGAUUAACCGACGCCAACGGGAGAAUGGCAG 35

RESULT 6
US-08-412-110-72
; Sequence 72, Application US/08412110
; Patent No. 5670637
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 344
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/412,110
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: June 10, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: June 11, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX01/C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-412-110-72

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

; FILING DATE: June 11, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX01/C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-412-110-72

Query Match 54.3%; Score 19; DB 1; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Oy 1 tcatgcattaccacccccctcgaggagcgagcag 35
:|||||:|||||:|||||:|||||
Db 1 UCAAGGAUUAACCGACGCCAACGGGAGAAUGGCAG 35

RESULT 7
US-08-412-110-94
; Sequence 94, Application US/08412110
; Patent No. 5670637
; GENERAL INFORMATION:
; APPLICANT: Larry Gold
; APPLICANT: Craig Tuerk
; TITLE OF INVENTION: Nucleic Acid Ligands
; NUMBER OF SEQUENCES: 344
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/412,110
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: June 10, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: June 11, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX01/C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-412-110-94

Query Match 54.3%; Score 19; DB 1; Length 41;

—

APPLICATION NUMBER: 08/412,110
FILING DATE: March 27, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/409,442
FILING DATE: March 24, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: June 10, 1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: June 11, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX01/C5
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-609A-72

Query Match 54.38; Score 19; DB 2; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgattaccacccctcgaggagcaggcag 35
:||| i::|| i|| i|| i|| i|| i|| i||
Db 1 UCAAGGAUUAACCGACGCGGAGGAGGAGGCGAG 35

RESULT 11

US-08-469-609A-94
Sequence 94, Application US/08469609A
Patent No. 5843653
GENERAL INFORMATION:
APPLICANT: Larry Gold
APPLICANT: Craig Tuerk
TITLE OF INVENTION: Nucleic Acid Ligands
NUMBER OF SEQUENCES: 374
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,609A
FILING DATE: June 6, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/428,964
FILING DATE: April 25, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/412,110
FILING DATE: March 27, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/409,442
FILING DATE: March 24, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: June 10, 1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/536,428
FILING DATE: June 11, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX01/C5
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-609A-94

Query Match 54.3%; Score 19; DB 2; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgattaccacccctcgaggagcaggcag 35
:||| i::|| i|| i|| i|| i|| i|| i||
Db 1 UCAAGGAUUAACCGACGCGGAGGAGGAGGCGAG 35

RESULT 12

US-09-143-190-72
Sequence 72, Application US/09143190
Patent No. 6110900
GENERAL INFORMATION:
APPLICANT: Larry Gold
APPLICANT: Craig Tuerk
TITLE OF INVENTION: Nucleic Acid Ligands
NUMBER OF SEQUENCES: 374
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/143,190
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/469,609
FILING DATE: June 6, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/428,964
FILING DATE: April 25, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/412,110
FILING DATE: March 27, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/409,442
FILING DATE: March 24, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: June 10, 1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: June 11, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX01/C6

us-09-904-420a-1_copy_866_900.rni

Thu Feb 21 07:25:45 2002

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-143-190-72

Query Match          54.3%; Score 19; DB 3; Length 41;
Best Local Similarity 62.9%; Pred. No. 19;
Matches 22; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 tcatgcattaccacccctcggagcaggcag 35
   :||| :||| :||| :||| :||| :||| :||| :|||
Db 1 UCAGGAUUUACCGACGCGCAACGGGAGAAUGGCAG 35

RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          53.1%; Score 18.6; DB 4; Length 4403765;
Best Local Similarity 84.0%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 ttaccacacccctcggagcagg 32
   ||||| ||| ||||| ||| |||
Db 4296944 TTACCACACTCCATCGGTGCACGG 4296920

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match 53.1%; Score 18.6; DB 4; Length 4411529;
Best Local Similarity 84.0%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 8 ttaccacacccccctcggagcagg 32
||||||| || |||| ||| ||
Db 4304689 TTACCCACCTCCATCGGTCACGG 4304665

Search completed: February 20, 2002, 14:32:12
Job time: 14391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 09:22:11 ; Search time 3793.25 Seconds
(without alignments)
99.150 Million cell updates/sec

Title: US-09-904-420A-1_COPY_866_900

Perfect score: 35

Sequence: 1 tcatgcattaccaccccccctcggagcagggcag 35

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	353	11	BG386035
2	33	94.3	273	10	AA311787
3	28.6	81.7	400	11	BG555601
4	27	77.1	649	11	BI154476
5	25.4	72.6	264	10	BB584341
6	22.4	64.0	951	11	BF795180
7	21.8	62.3	323	10	AW087211
8	21.8	62.3	919	11	BG190073
9	21.2	60.6	934	11	BF107882
10	21.2	60.6	947	10	AI377394
11	21.2	60.6	1718	11	BF183334
12	21	60.0	308	10	BB066565

c	13	21	60.0	329	10	AV047890
	14	21	60.0	445	11	BF606275
	15	20.8	59.4	247	10	AV205089
	16	20.8	59.4	254	10	BB245996
	17	20.8	59.4	256	10	BB381453
c	18	20.8	59.4	581	13	AQ480983
	19	20.8	59.4	674	13	AQ471695
	20	20.6	58.9	514	10	AW960157
c	21	20.6	58.9	657	11	BE177628
c	22	20.4	58.3	691	11	BG857928
c	23	20.4	58.3	733	13	A2962176
c	24	20.4	58.3	930	11	BE796914
c	25	20.2	57.7	218	11	BF407387
c	26	20.2	57.7	300	11	BF419958
c	27	20.2	57.7	314	10	AA875397
c	28	20.2	57.7	381	11	BG625362
c	29	20.2	57.7	498	10	AI091571
c	30	20.2	57.7	513	13	P52598
c	31	20.2	57.7	532	10	AI008676
c	32	20.2	57.7	533	10	AW913994
c	33	20.2	57.7	608	10	AL596682
c	34	20.2	57.7	948	11	BG025922
c	35	20	57.1	201	11	BF799331
c	36	20	57.1	233	10	BB544489
c	37	20	57.1	298	13	A2514882
c	38	20	57.1	381	13	A2049670
c	39	20	57.1	428	13	AZ049666
c	40	20	57.1	1065	13	CNS02J22
c	41	19.8	56.6	173	13	A2744475
c	42	19.8	56.6	194	11	BF935378
c	43	19.8	56.6	202	10	AV098892
c	44	19.8	56.6	207	11	BG383736
c	45	19.8	56.6	261	10	AW415192

ALIGNMENTS

RESULT 1

BG386035 353 bp mRNA EST 12-MAR-2001
LOCUS 602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5',
DEFINITION mRNA sequence.
ACCESSION BG386035
VERSION BG386035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 353)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1307 row: h column: 02
High quality sequence stop: 276.
Location/Qualifiers
1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"

FEATURES
source

/note="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGACGAG(c). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 80 a 111 c 122 g 40 t

ORIGIN

Query Match 100.0%; Score 35; DB 11; Length 353;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcattgcatlaccacccctcgagagagagcag 35
 Db 239 TCATGCATTACCCGCCCTCGGACAGAGCGCAG 273

RESULT 2
 AA311787 273 bp mRNA EST 19-APR-1997
 LOCUS EST182503 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
 DEFINITION similar to T-cell factor 1, A/B/C, mRNA sequence.
 AA311787
 VERSION AA311787.1 GI:1964114
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A., Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D., White O., Sutton G., Blake J.A., Brandon R.C., Man-Wei C., Clayton R.A., Cline R.R., Cotton M.D., Barle-Hughes J., Fine L.D., Fitzgerald L.M., Fitzhugh W.M., Fitchman J.L., Geoghegan N.S., Glodok A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr., Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M., Moreno-Palancas R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M., Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R., Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y., Bedarick D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J., Diike D., Feng D.F., Ferrie A., Fischer C., Hastings G.A., He M.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K., Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H., Raymond L., Wei Y.F., Wang J., Xu C., Yu G.L., Ruben S.M., Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C., Fraser C.M. and Venter J.C.

TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT Contact: Kerlavage, AR

CONTACT: Kerlavage, AR

Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavet@igf.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers

1..273
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):158998"
 /db_xref="taxon:9606"
 /clone_lib="Jurkat T-cells VI"
 /cell_type="T-lymphocyte"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 60 a 111 c 59 g 39 t 4 others

ORIGIN

Query Match 94.3%; Score 33; DB 10; Length 273;
 Best Local Similarity 94.3%; Pred. No. 0.072;
 Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tcattgcatlaccacccctcgagagagagcag 35
 Db 123 TCATGCATTACCCGCCCTCGGACAGAGCGCAG 157

RESULT 3
 BG655601 400 bp mRNA EST 05-JUL-2001
 LOCUS 1a84a02.x1 Melton Mouse E16 5 Pancreas Library M1621 Mus musculus
 DEFINITION cDNA 3' similar to SM:TCPLMOUSE 000417 T-CELL-SPECIFIC TRANSCRIPTION FACTOR 1; , mRNA sequence.
 BG655601
 VERSION BG655601.1 GI:13793010
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 400)

Melton D., Brown J., Kenty G., Permutt A., Lee C., Kaestner K., Lemisnka I., Searce M., Brestelli J., Gradwohl G., Clifton S., Hillier L., Marra M., Pape D., Wyle T., Martin J., Blistein A., Schmitt A., Theising B., Ritter E., Ronko J., Bennett J., Cardenas M., Gibbons M., McCann R., Cole R., Tsagarisvilli K., Williams T., Jackson Y. and Bowers Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Department of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
 Possible reversed clone: similarity on wrong strand
 Seq primer: -400p from Gibco
 High quality sequence stop: 394.

FEATURES
 source Location/Qualifiers

1..400
 /organism="Mus musculus"
 /strain="129"
 /db_xref="taxon:10090"
 /clone_lib="Melton Mouse E16 5 Pancreas Library M1621"
 /sex="Both"
 /issue_type="Total pancreas"
 /dev_stage="Embryonic day 16.5"
 /lab_host="TOP10"
 /note="Organ: Pancreas; Vector: pZFR0-2; Site_1: NotI; Site_2: XhoI; Library constructed using Superscript Plasmid Library Kit (Life Technologies). cDNA made by oligo-dt priming. XhoI site destroyed during cloning. Size-selected by column fractionation; average insert size 1.2kb. Primary library, unamplified."
 87 a 148 c 98 g 67 t

Query Match 81.7%; Score 28.6; DB 11; Length 400;

Best Local Similarity 88.6%; Pred. No. 2.4;
Matches 31; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 tcatgcatlaccacccccctcggaagcaggcag 35
 ||||| ||||| ||||| ||||| |||||
 Db 153 TCATGCCCTACCCACCGGCTCGGAGACAGGGCAG 18

	RESULT	4
B1154476		
LOCUS	649 bp	mRNA
DEFINITION	6022904810F1 NIH_GCAP_Mam3 Mus musculus	EST
		CDNA clone IMAGE:5034194 5'

ACCESSION	BI154476	GI:14614477
VERSION	BI154476.1	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE 1 (bases 1 to 649)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Tissue procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MSC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMU at:
<http://image.lmu.gov>
 Plate: LM11095 row: d column: 03
 High quality sequence stop: 647.

FEATURES	Location/Qualifiers
source	1. .649

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="5034194"
/clone_id="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10b"
/note="Organ: mammary; Vector: pCMV-SpOrb6; Site_1: Noli; Site_2: Sali; Cloned unidirectionally. Primer: Oligo dT. Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennigausen/Chu-Xia Deng, NIH Reference sample. Transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NCI_CGAP Library."

```

BASE COUNT	157 a	204 c	167 g	121 t
ORIGIN				

Query Match:	77.1%	Score 27:	DB 11:	Length 649:
Best Local Similarity:	85.7%	Pred. No. 8.7:		
Matches 30:	Conservative 0:	Mismatches 5:	Indels 0:	Gaps 0:

QY	1	lcaigcattaccacccccctcgagagcagggcag	35
Db	341	TCATGCTTCACCCACCGGCTCCGGAGGAGGCGAG	375

RESULT	5				
BB584.341					
LOCUS	264 bp	mrna	EST	30-NOV-2000	
DEFINITION	BB584.341	BB584.341	Riken full-length enriched, adult male epididymis Mus		
	musculus cDNA clone 9230105F22.5',		mrna sequence.		

ACCESSION	BB584341
VERSION	BB584341.1
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE AUTHORS

TITLE
JOURNAL
COMMENT

Eikaiyotae, Metazoa: Chordata: Craniata: Euteleostomi, Mammalia, Eutheria: Rodentia; Sciurognathi, Muridae; Mus, 1 (basses 1 to 264)

Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arikawa, T., Carlinici, P., Hasegawa, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodyama, Y., Imotoi, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Komatsu, K., Kusabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sekai, K., Sasai, D., Sato, C., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toyota, T., Watanishi, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.

Riken Mouse ESTs (Alizawa, K. et al. 2000)

Unpublished (2000)

Contact: Yoshihide Hayashizaki
Riken Developmental Biology Research Group, RIKEN Genomic

Research Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome_research@ken.go.jp,
URL: <http://genome.igsc.iken.go.jp/>
Carninci, P., Nishiyasu, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,
, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Thermolabile and thermocyclable restriction enzymes by
thermostabilization and thermocyclization for the synthesis of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Katsuno, T., Akiyama, K., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (3), 463-470 (1999)
Carninci, P., and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. METEORUS ENZYMOLOGY 19-44 (1999)

please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source

```

location/Qualifiers
1..264
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9230105622"
/clone_lib="RIKEN full-length enriched, adult male
epididymus"
/sex="male"
/tissue_type="epididymus"
/dev_stage="adult"
/lab_host="PH10B"
/insert="1. Salt. Site 3. BamHI. cDNA library W

```

70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Noted=Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genetic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGCAGACGAGAGATTCACAAAGCTCTTTTTPTTTTTNN 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGCAGACGATTCGCGATTAATTAAATTCAGCCCCCCCC 3'. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmidscript KS(+) after bulk excision from Lambda PLIC 1. Cloning sites, 5' end: SalI; 3' end: BamHI.

BASE COUNTRY
ORIGIN

BASE COUNT	54 a	67 c	76 g	67 l
Query Match	72.6%	Score 25.4	DB 10	length 264

Best Local Similarity 82.9%: Pred. No. 31;
Matches 29; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 tcatgcatcaccaccctcgagcagcagcag 35
Db 173 TCATGCCCTACCCACCGCTCCGAGCAGCGGAG 207

RESULT 6
BF795180 951 bp mRNA EST 12-JAN-2001
LOCUS 602296329F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4339687 5',
DEFINITION mRNA sequence.
ACCESSION BF795180
VERSION BF795180.1 GI:12100234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 951)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.llnl.gov>
Plate: LHAM9950 row: j column: 08
High quality sequence stop: 684.

FEATURES

SOURCE

1..951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4339687"
/clone_1lb="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lymph; Vector: pCMV-SPORT6; Site: j; NCI;
Site: 2; Salt: Cloned unidirectionally; oligo-dT primed.
Average insert size 1,867 bp. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 219 a 267 c 282 g 183 t
ORIGIN

Query Match 64.0%: Score 22.4; DB 11; Length 951;
Best Local Similarity 81.2%: Pred. No. 3 4e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 tgcattaccacccctcgagcagcagcagcag 35
Db 770 TCATTCACATCCCTCTCAGAGCAGCGGAG 739

RESULT 7
AM087211 323 bp mRNA EST 15-OCT-1999
LOCUS xb36e08.x1 NCI_CGAP_Lu31 Homo sapiens cDNA clone IMAGE:2578406 3',
DEFINITION mRNA sequence.
ACCESSION AM087211
VERSION AM087211.1 GI:6043016
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 323)
AUTHORS NCI-CGAP <http://www.nci.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: ATCC cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/dbip/image/image.html
Seq primer: -40up from gldco
High quality sequence stop: 311.

FEATURES

SOURCE

1..323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2578406"
/clone_1lb="NCI_CGAP_Lu31"
/sex="male"
/dev_stage="fetal, 14 wk post-conception"
/lab_host="DH10B"
/note="Organ: lung, cell line: Vector: pCMV-SPORT6;
Site: 1; EcorV; Site: 2; NotI; Cloned unidirectionally, no
5' adaptor. Primer: oligo dT. Full-length library
constructed by Life Technologies."

BASE COUNT 124 a 92 c 73 g 34 t
ORIGIN

Query Match 62.3%: Score 21.8; DB 10; Length 323;
Best Local Similarity 78.8%: Pred. No. 5.5e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 catgcatcaccacccctcgagcagcagcag 34
Db 248 CAGGCTTACCCCGCCCGGCGGAGGAGGAA 280

RESULT 8
BG190073 919 bp mRNA EST 21-APR-2001
LOCUS RSR9129 Altherys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG190073
ACCESSION BG190073
VERSION BG190073.1 GI:13711760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 919)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, D.,
Lerner, L., Krishoc, D., McElligott, K., Clark, S., Mays, K., Smith, E.,
Veloso, N., Hess, J., Colhren, K., Lo, K., Offenbacher, J., Panzig, J.,
and Ducar, M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
CONTACT: Scott J. Cain
Altherys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@altherys.com
High quality sequence stop: 438.

TITLE
JOURNAL
COMMENT

FEATURES

SOURCE

1..919
/organism="Homo sapiens"
/db_xref="taxon:9606"

COMMENT	FEATURES
<p>Contact: Yoshihiro Nagashizaki, Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URL: http://genome.res.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki , N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Thermostabilization and thermocatalysis of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki , Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.</p>	<p>Location/Qualifiers</p> <p>1..308</p> <p>organism="Mus musculus"</p> <p>strain="C57BL/6J"</p> <p>db_xref="taxon:10090"</p> <p>clone="80304536C20"</p> <p>clone_id="RIKEN full-length enriched, 15 days embryo male testis"</p> <p>sex="male"</p> <p>tissue_type="testis"</p> <p>dev_stage="15 days embryo"</p> <p>lab_host="DH10B"</p> <p>note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGCAGAGGAGCATCCAGACGCTTTTGTGTTTGTGTTTNN 3']. cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rat = 185.0 Second strand cDNA was prepared with the primer adapter of sequence [5', GAGCAGAGGATTCGACGATTAATTAATTAATTCACCCGCCCG 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC19 after bulk excision from Lambda phage. Cloning sites, 5' end: SalI; 3' end: BamHI"</p>
<p>BASE COUNT</p> <p>74 a 71 c 49 g 114 t</p> <p>ORIGIN</p>	
<p>Query Match</p> <p>Best Local Similarity 82.8%; Pred. No. 1e+03;</p> <p>Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;</p>	
<p>6 cattaccacacccclcgagagcagcga 34</p> <p> </p> <p>19 CATACCCACCCCTTCAGCAGCAGCA 47</p>	
<p>RESULT 13</p> <p>AV047890</p> <p>LOCUS</p> <p>AV047890 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone</p> <p>DEFINITION</p> <p>1700071101, mRNA sequence.</p> <p>AV047890</p> <p>AV047890.2 GI:4867555</p>	<p>329 bp</p> <p>mRNA</p> <p>EST</p> <p>23-NOV-1999</p>

KEYWORDS EST.
house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 329)
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Tagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@r.riken.go.jp
Thermolabile and thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome-rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1..329
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1700071L01"
/clone_1lb="Mus musculus adult C57BL/6J testis"
/sex="male"
/tissue="testis"
/dev_stage="adult"
79 g 61 t
BASE COUNT 97 a 92 c 61 t
ORIGIN

Query Match 60.0%; Score 21; DB 10; Length 329;

Best Local Similarity 82.8%; Pred. No. 1e+03; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 catgcatlaccacccccctcgagagacg 30
||||| ||||||| |||||||
Db 54 CATGCAGACCCACCCACCGGAGCTG 82

RESULT 14
B606275 445 bp mRNA EST 25-APR-2001
LOCUS B606275
DEFINITION B606275
ACCESSION B606275
VERSION B606275.1 GI:11706718
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 445)
AUTHORS Smith, T.P., L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Cass, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G., Petrea, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J., and Keene, J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@ma1.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACCTATGACCT
BACKWARD: GTTTCACGTCACGACG
Plate: 60 row: A column: 12
Seq primer: ATTTCAGTCACCTATGACG
Location/Qualifiers

FEATURES

Location/Qualifiers
1..445
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="MARC 3BOV"
/tissue="pooled"
/lab_host="DH10B"
/note="Vector: PCWV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
78 a 121 c 141 g 105 t
BASE COUNT 78 a 121 c 141 g 105 t
ORIGIN

Query Match 60.0%; Score 21; DB 11; Length 445;

Best Local Similarity 82.8%; Pred. No. 1e+03; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 gcatlaccacccccctcgagagacg 33
||||| ||||||| |||||||
Db 235 GCCTTACTACCCCTCCCTCGAGGAC 207

RESULT 15
AV205089 247 bp mRNA EST 30-OCT-1999
LOCUS AV205089
DEFINITION AV205089 RIKEN full-length enriched, adult male testis Mus musculus
ACCESSION AV205089
VERSION AV205089.1 GI:6145942
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 247)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadoh, K., Kadawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
TITLE RIKEN Mouse ESTs (Konno, H., et al. 1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuho-cho, Tsukuba-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@r.riken.go.jp,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:24:46 ; Search time 1729.86 Seconds
(without alignments)
333.785 Million cell updates/sec

Title: US-09-904-420a-1_COPY_851_885

Perfect score: 1 cgcgcctcaatcgtcgcacgaattaccaccccc 35

Sequence: 1 cgcgcctcaatcgtcgcacgaattaccaccccc 35

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_hlg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_to: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htgo_hum: *
31: em_htgo_iny: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_iny: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	1115	9	HSTCF1D
2	35	100.0	1165	9	HSTCF1A
3	35	100.0	1254	9	HSTCF1B
4	35	100.0	2814	9	HSTCF1C
5	35	100.0	2855	9	HSTCF1D
6	35	100.0	2910	9	HSTCF1E
7	35	100.0	84544	2	AC009012
8	35	100.0	86914	2	AC011336
9	35	100.0	159420	2	AC009017
10	35	100.0	200831	2	AC008608
11	27	77.1	11790	10	MMTCSTF
12	27	77.1	162496	2	AC044846
13	23.8	68.0	150434	2	AL138784
14	23.8	68.0	151349	2	AL353647
15	23.8	68.0	155332	2	AL365439
16	23.8	68.0	160511	2	AC022786
17	22.4	64.0	128155	9	HS246928
18	21.8	62.3	171915	9	HS75613
19	21.8	62.3	340000	9	HS21C009
20	21.8	61.1	7982	2	AC014670
21	21.4	61.1	20950	2	AC003242
22	21.4	61.1	31804	2	AC014202
23	21.4	61.1	81602	3	AC004349
24	21.4	61.1	154080	2	AC024416
25	21.4	61.1	162796	2	AC011927
26	21.4	61.1	191142	2	AL596129
27	21.4	61.1	214713	3	AE003445
28	21.4	61.1	313965	3	AE003445
29	21.2	60.6	42546	2	AC011555
30	21.2	60.6	179823	2	AC092067
31	21.2	60.6	188481	9	AL138836
32	21.2	60.6	204037	2	AC015881
33	21.2	60.6	183256	2	AC091714
34	21.2	60.6	183256	2	AP000345
35	20.8	59.4	151600	9	AP000346
36	20.8	59.4	151600	9	AP000346
37	20.8	59.4	157677	2	AP002470
38	20.8	59.4	165330	10	AC087233
39	20.8	59.4	166063	9	AP002515
40	20.8	59.4	168108	2	AC084775
41	20.8	59.4	175464	2	AP002887
42	20.8	59.4	205347	2	AL392111
43	20.6	58.9	2676	1	ECMSYB
44	20.6	58.9	10980	1	AE000206
45	20.6	58.9	12029	1	AE005316

ALIGNMENTS

RESULT 1 HSTCF1D HSTCF1D 1115 bp mRNA PRI 09-JAN-1995

LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form D.

DEFINITION Z47361.1 GI:619881

ACCESSION Z47361.1 GI:619881

VERSION splice form D; T cell factor 1.

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1115)

MAVER K., Wolff E., Clevers H. and Ballhausen W.G.

The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon

JOURNAL, Unpublished

2 (bases 1 to 1115)

van de Wetering, M., Oosterwegel, M., Holstege, F., Doyes, D.,

Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.

TITLE The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
JOURNAL J. Biol. Chem. 267 (12), 8530-8536 (1992)
MEDLINE 92235082
REFERENCE 3 (bases 1 to 1115)
AUTHORS Ballhausen, W.G.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES
source
Location/Qualifiers
1..1115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1..1115
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1..1115
/gene="TCF-1"
2..754
/gene="TCF-1"
/standard_name="T cell factor 1 splice form D"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form D"
/protein_id="CAA87439.1"
/db_xref="GI:619882"
/translation="MYKRTVSAFNLIMHPPPSGAGQHPQPPPLHKANQPHGVQ
LSLVEHNSPHPPADISQKQVHRLQTPDLSGFYSLTSGMGLPHVTSWTFHS
LMGSGVGHAAIPPAIVPPSGKQELQPPRNILQTAESKAEKAPN1KKPLNA
FMLYKEMRAKVIACETLKESAA1NO1LRRHNAISREDAKYELAKRERQJLMOLY
PGMSARDNIGKKRRSRERKQSTTDSNLSHS"
mat_peptide
731..751
/gene="TCF-1"
/product="alternative ORF specific for TCF-1D"
BASE COUNT 272 a 407 c 265 g 171 t
ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 1115;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgcctcaatcgtcatgcataccaccccc 35
|||||
Db 24 ccgcctcaatcgtcatgcataccaccccc 58

RESULT 2
HSTCF1E 1165 bp mRNA PRI 09-JAN-1995
LOCUS H.sapiens TCF-1 mRNA for T cell factor 1 splice form E.
DEFINITION 247362
ACCESSION 247362.1 GI:619883
VERSION splice form E; T cell factor 1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1165)
AUTHORS Mayer, K., Wolff, E., Clevers, H. and Ballhausen, W.G.
TITLE The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon
JOURNAL IXA
REFERENCE 2 (bases 1 to 1165)
AUTHORS van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D.,

TITLE Suikerbuijk, R., Geurts van Kessel, A. and Clevers, H.
JOURNAL The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
MEDLINE 92235082
REFERENCE 3 (bases 1 to 1165)
AUTHORS Ballhausen, W.G.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Ballhausen W. G., Institut fuer Humangenetik der Universitaet, Universitaet Erlangen-Nuernberg, Schwabachanlage 10, Erlangen, Germany, D-91054

FEATURES
source
Location/Qualifiers
1..1165
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lymphoma"
/cell_type="T-lymphocyte"
/cell_line="Jurkat"
1..1165
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/function="high mobility group box transcription factor"
/evidence=experimental
1..1165
/gene="TCF-1"
2..1144
/gene="TCF-1"
/standard_name="T cell factor 1 splice form E"
/citation=[1]
/codon_start=1
/product="T cell factor 1 splice form E"
/protein_id="CAA87440.1"
/db_xref="GI:619884"
/translation="MYKRTVSAFNLIMHPPPSGAGQHPQPPPLHKANQPHGVQ
LSLVEHNSPHPPADISQKQVHRLQTPDLSGFYSLTSGMGLPHVTSWTFHS
LMGSGVGHAAIPPAIVPPSGKQELQPPRNILQTAESKAEKAPN1KKPLNA
FMLYKEMRAKVIACETLKESAA1NO1LRRHNAISREDAKYELAKRERQJLMOLY
PGMSARDNIGKKRRSRERKQSTTDSNLSHS"
mat_peptide
803..1141
/gene="TCF-1"
/product="alternative ORF specific for TCF-1E"
BASE COUNT 279 a 422 c 286 g 178 t
ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 1165;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccgcctcaatcgtcatgcataccaccccc 35
|||||
Db 24 ccgcctcaatcgtcatgcataccaccccc 58

RESULT 3
HSTCF1A 1254 bp mRNA PRI 14-JUN-1991
LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form A).
DEFINITION X59869 X55327
ACCESSION X59869.1 GI:36785
VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1254)
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

REFERENCE The Netherlands
2 (bases 1 to 1254)
AUTHORS van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H.
TITLE Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box
JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 9114695
COMMENT See also X59869-X59871.
FEATURES
Source
Location/Qualifiers
1..1254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat"
/clone_lib="CDNA"
/clone_id="TCF-1a"
1..1254
/gene="TCF-1"
/evidence="experimental"
1..1254
/gene="TCF-1"
80..889
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1, splice form A"
/protein_id="CAA42526.1"
/db_xref="GI:36786"
/db_xref="SWISS-PROT:P36402"
/translation="MYKETYSAFNLMLHYPPPSGAGOHPOPOPPLKANKNPPHGVPO
LSLYEHNSPHPTAPADISOKVHRPIQTPDLSGYSLTSGMGLPHVTSMPTHS
LMLSGVPGHPAAIHPHAIYPPSGKOELOPDRNLKTQAEKAEKAKPTTKPLNA
FMLYMKEMRAKVIAECTLKESAINOILGRMHALSREOKAYEIAKREOLHMOYL
PGWSARDNYGKKRRSRKREKHQESTTETMPRELKDGNGOESLSMSSSSPA"
misc_feature
539..769
/gene="TCF-1"
/note="HMG box"
810
/gene="TCF-1"
/note="alternative splice site"
BASE COUNT 320 a 391 c 335 g 208 t
ORIGIN
Query Match 100.0%; Score 35; DB 9; Length 1254;
Best local Similarity 100.0%; Pred. No. 0.00029;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
LOCUS HSTCF1B 2814 bp mRNA PRI 17-JUN-1991
DEFINITION Human TCF-1 mRNA for T cell factor 1 (splice form B).
ACCESSION X59870.X5329
VERSION X59870.1 GI:36787
KEYWORDS DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS van de Wetering, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical
Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,
The Netherlands
2 (bases 1 to 2814)
REFERENCE
AUTHORS van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H.
TITLE Identification and cloning of TCF-1, a T lymphocyte-specific
transcription factor containing a sequence-specific HMG box

JOURNAL EMBO J. 10 (1), 123-132 (1991)
MEDLINE 9114695
COMMENT See also X59869-X59871.
FEATURES
Source
Location/Qualifiers
1..2814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T lymphocyte"
/cell_line="Jurkat and HEB-ALL"
/clone_lib="CDNA"
1..2814
/gene="TCF-1"
/evidence="experimental"
1..2814
/gene="TCF-1"
80..889
/gene="TCF-1"
/codon_start=1
/product="T cell factor 1 splice form B"
/protein_id="CAA42527.1"
/db_xref="GI:36788"
/db_xref="SWISS-PROT:P36402"
/translation="MYKETYSAFNLMLHYPPPSGAGOHPOPOPPLKANKNPPHGVPO
LSLYEHNSPHPTAPADISOKVHRPIQTPDLSGYSLTSGMGLPHVTSMPTHS
LMLSGVPGHPAAIHPHAIYPPSGKOELOPDRNLKTQAEKAEKAKPTTKPLNA
FMLYMKEMRAKVIAECTLKESAINOILGRMHALSREOKAYEIAKREOLHMOYL
PGWSARDNYGKKRRSRKREKHQESTTETMPRELKDGNGOESLSMSSSSPA"
misc_feature
539..769
/gene="TCF-1"
/note="HMG box"
810
/gene="TCF-1"
/note="alternative splice site"
BASE COUNT 763 a 831 c 648 g 572 t
ORIGIN
Query Match 100.0%; Score 35; DB 9; Length 2814;
Best local Similarity 100.0%; Pred. No. 0.00028;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
LOCUS HSTCF1G 2855 bp DNA PRI 30-SEP-1999
DEFINITION Homo sapiens TCF-1 gene.
ACCESSION X63901
VERSION X63901.1 GI:36791
KEYWORDS T-cell transcription factor; transcription factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS van de Wetering, M., Oosterwegel, M., Holsiege, F., Dooijes, D.,
Suijkerbuijk, R., Geurts van Kessel, A. and Clevers, H.
TITLE The human T cell transcription factor-1 gene. Structure,
localization, and promoter characterization
J. Biol. Chem. 267 (12), 8530-8536 (1992)
JOURNAL 92235082
MEDLINE
REFERENCE
AUTHORS van de Wetering, M., Castrop, D., Korfink, V. and Clevers, H.
TITLE Extensive alternative splicing and dual promoter usage generate
TCF-1 protein isoforms with differential transcription control
properties
Mol. Cell. Biol. 16 (3), 745-752 (1996)
JOURNAL 96182076
MEDLINE
REFERENCE
AUTHORS van de Wetering, M.L.

TITLE Direct Submission
JOURNAL Submitted (07-JAN-1992) M.L. Van De Wetering, Department of Immunology, P.O. Box 8550, 3508 GA Utrecht, THE NETHERLANDS

FEATURES
Source Location/Qualifiers

1..2855

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/cell_type="T-lymphocyte"

/clone="phi-TCF-[5,8,11]"

/map="5q31.1"

829..924

/gene="TCF-1"

/number=1

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

/codon_start=1

/protein_id="CAB56795.1"

/db_xref="GI:6006565"

/translation="MYKETYVSAFNLIMHYPPSGAGOHPOPPPLKANKOPPHGVPO

LSLTFHNSPHPTPAPADISOKOVHRLPTPLDSGYSLTSGSMGQLPPIVSWFTPHS

LMISGVGHPAAIPHPAIVPESGKLOPFDNNLKTQAESKAEREAKKPTIKKPLNA

FMLYMKEMRAKVIACETLKESAAINOLGRMHALSREDAKYVELAKKEQOLHMQLT

PGMSARDNYGKKRKRREKHQESTTTPNMPRELKDNQGESLSMSSSSSPA"

join(829..924,970..1074,1120..1209,1255..1374,1420..1581,

1691..1798,1843..1890,2054..2134)

/gene="TCF-1"

925..970

/number=1

970..1074

/gene="TCF-1"

/number=2

1075..1119

/number=2

1120..1209

/gene="TCF-1"

/number=3

1210..1254

/number=3

1255..1374

/gene="TCF-1"

/number=4

1375..1419

/number=4

1420..1581

/gene="TCF-1"

/number=5

1582..1690

/number=5

1691..1798

/gene="TCF-1"

/number=6

1799..1842

/number=6

1843..1890

/gene="TCF-1"

/number=7

1891..2053

/number=1

2054..2134

/gene="TCF-1"

/number=8

BASE COUNT 592 a 929 c 771 g 562 t 1 others

ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 2855;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccttcaatctgctcatgattaccaccccc 35
|||||

Db 851 CCGCTTCAATCTGCTCATGATTACCCACCCCC 885

RESULT 6 HSTCF1C 2910 bp mRNA PRI 17-JUN-1991

LOCUS Human TCF-1 mRNA for T cell factor 1 (splice form C).

DEFINITION X59871.X55328

ACCESSION X59871.1 GI:36789

VERSION DNA-binding protein; HMG box; T cell factor 1; TCF-1 gene.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2910)

van de Wetering, M.

Direct Submission

Submitted (28-MAY-1991) M. Van De Wetering, Dept of Clinical

Immunology, University Hospital, P.O. Box 85500, 3508 GA Utrecht,

The Netherlands

2 (bases 1 to 2910)

van de Wetering, M., Oosterwegel, M., Doelies, D. and Clevers, H.

Identification and cloning of TCF-1, a T lymphocyte-specific

transcription factor containing a sequence-specific HMG box

EMBO J. 10 (1), 123-132 (1991)

91114695

See also X59869-X59871.

FEATURES

Location/Qualifiers

1..2910

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="T lymphocyte"

/clone_lib="cDNA"

1..2910

/gene="TCF-1"

/evidence="experimental"

1..2910

/gene="TCF-1"

80..886

/gene="TCF-1"

/codon_start=1

/product="T cell factor 1, splice form C"

/protein_id="CAA42528.1"

/db_xref="GI:36790"

/db_xref="SWISS-PROT:P36402"

/translation="MYKETYVSAFNLIMHYPPSGAGOHPOPPPLKANKOPPHGVPO

LSLTFHNSPHPTPAPADISOKOVHRLPTPLDSGYSLTSGSMGQLPPIVSWFTPHS

LMISGVGHPAAIPHPAIVPESGKLOPFDNNLKTQAESKAEREAKKPTIKKPLNA

FMLYMKEMRAKVIACETLKESAAINOLGRMHALSREDAKYVELAKKEQOLHMQLT

PGMSARDNYGKKRKRREKHQESTTTPNMPRELKDNQGESLSMSSSSSPA"

539..769

/gene="TCF-1"

/note="HMG box"

810

/gene="TCF-1"

/note="alternative splice site"

BASE COUNT 782 a 860 c 671 g 597 t

ORIGIN

Query Match 100.0%; Score 35; DB 9; Length 2910;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccttcaatctgctcatgattaccaccccc 35
|||||

Db 102 CCGCTTCAATCTGCTCATGATTACCCACCCCC 136

RESULT 7

AC009012/c 84544 bp DNA HMG 19-APR-2001

DEFINITION Homo sapiens chromosome 5 clone XXP1-360D11, WORKING DRAFT
SEQUENCE, 7 ordered pieces.
ACCESSION AC009012
VERSION AC009012.3 GI:12830104
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 84544)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 84544)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 14, 2001 this sequence version replaced gi:7454202.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1189075, H71
Center clone name: XXP1-360D11

Summary Statistics
Consensus quality: 78189 bases at least Q40
Consensus quality: 82001 bases at least Q30
Consensus quality: 83201 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 83944; sum-of-contigs estimation
Quality coverage: 6.34 in Q20 bases; pulse field gel estimation
Quality coverage: 6.04 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2979: contig of 2979 bp in length
* 2980 34576: gap of unknown length
* 3080 34576: contig of 31497 bp in length
* 34577 34676: gap of unknown length
* 34677 59262: contig of 24386 bp in length
* 59263 59362: gap of unknown length
* 59363 76010: contig of 16648 bp in length
* 76011 76110: gap of unknown length
* 76111 79613: contig of 3503 bp in length
* 79614 79713: gap of unknown length
* 79714 80648: contig of 936 bp in length
* 80650 80749: gap of unknown length
* 80750 84544: contig of 3795 bp in length.
Location/Qualifiers
1. 84544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXP1-360D11"
BASE COUNT 19558 a 21740 c 22203 g 20376 t 667 others
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 84544;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 ccgcctcaatctgctcatgataccaccccc 35

|||||
Db 38717 CCgcCTTCAATCTGCTCATGATACCACCCCC 38683
RESULT 8
LOCUS AC011336 86914 bp DNA HTG 23-APR-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-250113, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC011336.4 GI:9256274
VERSION AC011336
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 86914)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 86914)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 18, 2000 this sequence version replaced gi:7710593.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 281384, H247
Center clone name: CIT-HSPC_250113

Summary Statistics
Consensus quality: 82614 bases at least Q40
Consensus quality: 85602 bases at least Q30
Consensus quality: 86382 bases at least Q20
Estimated insert size: 89000; pulse field gel estimation
Estimated insert size: 86914; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 6.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86914: contig of 86914 bp in length.
Location/Qualifiers
1. 86914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-250113"
BASE COUNT 20745 a 22135 c 22612 g 21422 t
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 86914;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 ccgcctcaatctgctcatgataccaccccc 35
|||||
Db 45292 CCgcCTTCAATCTGCTCATGATACCACCCCC 45288

RESULT 9
AC009017/c 159420 bp DNA HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone XXp1-929G6, WORKING DRAFT SEQUENCE,
DEFINITION 26 unordered pieces.
ACCESSION AC009017 GI:13699469
VERSION AC009017.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 159420)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159420)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:11178048.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1189133, H51
Center clone name: XXp1-929G6

Summary Statistics
Consensus quality: 142845 bases at least Q40
Consensus quality: 150910 bases at least Q30
Consensus quality: 152756 bases at least Q20
Estimated insert size: 80000; pulse field gel estimation
Estimated insert size: 156920; sum-of-coverage estimation
Quality coverage: 8.41 in Q20 bases; sum-of-coverage estimation
Quality coverage: 4.29 in Q20 bases; sum-of-coverage estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1423: contig of 1423 bp in length
1
1424 1523: gap of unknown length
1524 3066: contig of 1543 bp in length
3067 3166: gap of unknown length
3167 5629: contig of 2463 bp in length
5630 5729: gap of unknown length
5730 6997: contig of 1268 bp in length
6998 7098: gap of unknown length
7099 9214: contig of 2117 bp in length
9215 9314: gap of unknown length
9316 11813: contig of 2499 bp in length
11814 11913: gap of unknown length
11914 14718: contig of 2805 bp in length
14719 14818: gap of unknown length
14819 17174: contig of 2356 bp in length
17175 17274: gap of unknown length
17275 19710: contig of 2436 bp in length
19711 19810: gap of unknown length
19811 21678: contig of 1868 bp in length
21679 21778: gap of unknown length
21779 26021: contig of 4243 bp in length
26022 26121: gap of unknown length
26122 29609: contig of 3488 bp in length
29610 29710: gap of unknown length
29711 33197: contig of 3487 bp in length
33197 33296: gap of unknown length

33297 36995: contig of 3699 bp in length
36996 37095: gap of unknown length
37096 40646: contig of 3551 bp in length
40647 40746: gap of unknown length
40747 45044: contig of 4298 bp in length
45045 45144: gap of unknown length
45145 52148: contig of 7004 bp in length
52149 52248: gap of unknown length
52249 57686: contig of 5438 bp in length
57687 57786: gap of unknown length
57787 66113: contig of 8327 bp in length
66114 66213: gap of unknown length
66214 75538: contig of 9325 bp in length
75539 75638: gap of unknown length
75639 82249: contig of 6611 bp in length
82249 82350: gap of unknown length
82350 95465: contig of 13116 bp in length
95465 95566: gap of unknown length
95566 110942: contig of 15377 bp in length
110943 111042: gap of unknown length
111043 127290: contig of 16248 bp in length
127290 127390: gap of unknown length
127391 143681: contig of 16291 bp in length
143682 143781: gap of unknown length
143782 159420: contig of 15639 bp in length.

FEATURES
source 1..159420
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXp1-929G6"
BASE COUNT 33367 a 38414 c 39218 g 39915 t 2506 others
ORIGIN
Query Match 100.0%; Score 35; DB 2; Length 159420;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 cgcgcctcaatcgtcctatgcatccaccacc 35
Db 56846 CCCGCTTCATCTGCTCATGCAATFACCAACCC 56812

RESULT 10
AC008608 200831 bp DNA HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTB-113120, WORKING DRAFT SEQUENCE,
DEFINITION 29 unordered pieces.
ACCESSION AC008608
VERSION AC008608.5 GI:13699394
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 200831)
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE 2 (bases 1 to 200831)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7708985.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 107808, H233

Center clone name: CIT9785KB_113120

Summary Statistics

Summary statistics	17473 bases at least Q40
Consensus quality:	186693 bases at least Q30
Consensus quality:	190547 bases at least Q20
Estimated insert size:	165000: sum-of-coverage estimation
Estimated insert size:	198031: pulse-field gel estimation
Quality coverage:	7.77 in Q20 bases; sum-of-coverage estimation.
Quality coverage:	6.47 in Q20 bases; sum-of-coverage estimation.

NOTE: This is a working draft sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1115:	contig of 1115 bp in length
1116	1215:	gap of unknown length
1216	2567:	contig of 1352 bp in length
2568	2667:	gap of unknown length
2668	3745:	contig of 1078 bp in length
3746	3845:	gap of unknown length
3846	4920:	contig of 1075 bp in length
4921	5020:	gap of unknown length
5021	6044:	contig of 1024 bp in length
6045	6144:	gap of unknown length
6146	7285:	contig of 1141 bp in length
7286	7385:	gap of unknown length
7386	8659:	contig of 1274 bp in length
8660	8759:	gap of unknown length
8760	9804:	contig of 1045 bp in length
9805	9904:	gap of unknown length
9905	11696:	contig of 1792 bp in length
11697	11796:	gap of unknown length
11797	12841:	contig of 1045 bp in length
12842	12941:	gap of unknown length
12942	15043:	contig of 2102 bp in length
15044	15143:	gap of unknown length
15144	17289:	contig of 2146 bp in length
17290	17389:	gap of unknown length
17390	20232:	contig of 2843 bp in length
20232	20232:	gap of unknown length
20233	21884:	contig of 1552 bp in length
20333	21884:	contig of 1552 bp in length
21885	21984:	gap of unknown length
21985	24278:	contig of 2294 bp in length
24279	24379:	gap of unknown length
24379	28103:	contig of 3725 bp in length
28104	28203:	gap of unknown length
28204	30867:	contig of 2664 bp in length
30868	30967:	gap of unknown length
30968	34914:	contig of 3947 bp in length
34915	35014:	gap of unknown length
35015	36030:	contig of 1016 bp in length
36031	36130:	gap of unknown length
36131	38852:	contig of 2722 bp in length
38853	38952:	gap of unknown length
38953	41871:	contig of 2919 bp in length
41872	41971:	gap of unknown length
41972	47080:	contig of 5109 bp in length
47081	47180:	gap of unknown length
47181	49750:	contig of 2570 bp in length
49751	49850:	gap of unknown length
49851	57118:	contig of 7268 bp in length
57119	57218:	gap of unknown length
57219	66046:	contig of 8828 bp in length
66047	66146:	gap of unknown length
66147	81046:	contig of 14900 bp in length
81047	81146:	gap of unknown length
81147	101650:	contig of 20504 bp in length
101651	101750:	gap of unknown length
101751	137108:	contig of 33558 bp in length

FEATURES
source

	Location/Qualifiers
* 137109	137208: gap of unknown length
* 137209	200831: contig of 63623 bp in length.
	1..200831

[illegible]

Query Match	100.0%;	Score 35;	DB 2;	Length 200831;
Best Local Similarity	100.0%;	Pred. No. 0.00026;		
Matches 35; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

```
Oy      1 ccgccttcacatcgtgcacatgcatlaccacaccccc 35
          |||||
Db 184008 ccgccttcaatctgtctcatgtcatTACCCGACCCCC 184042
```

```

      RESULT 11
      MMTCTSF

```

DEFINITION Mouse mRNA for T-cell specific transcription factor.

```

VERSION      X61385.1      GI:54760
KEYWORDS     HMG box; T-cell specific transcription factor; transcription

```

SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE
1 (bases 1 to 1790)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

TITLE Direct Submission
JOURNAL Submitted (12-AUG-1991) M.A. Osterwegel, Dept of Clin Immunol,

NETHERLANDS
REFERENCE 2 (bases 1 to 1790)

Winoto, A., Georgopoulos, K., Meijlink, F. and Clevers, H.
Cloning of murine TCF-1, a T cell-specific transcription factor

J. Exp. Med. 173 (5), 1133-1142 (1991)

MEDLINE	11211020
FEATURES	location/Qualifiers
SOURCE	1 1790

```

/organism="Mus musculus
/isolate="M2a/M5"
/ab_xref="taxon:10080"

```

```

/issue_type="crim"
/cell_type="cell"

```

mrna	1. .1790
cds	190. .1101

```
/product="T-cell sp
/protein_id="CAA436
```

```

/db_xref="MGD:MCI:9
/db_xref="SWISS-PRO

```

QLSPLYEHFSSPHPTPAPA
PPIYPISPSGCGYROHPAP

NOEED I DNNEN I QAL I K
NÖILGRWHALSREEQAKY
TCCRBNAECTYDERAANBA

```
misc_feature      489. 381
/note="ACT splice/
```


[illegible]

CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Feb 9, 2001 this sequence version replaced gi:12227304.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: EM, EMBL; SW,
 SWISSPROT; Tr, TREMBL; WP, WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP5-1102M4 is from the library RPI-5 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTORE: pcVPAC2
 This sequence is the entire insert of clone RP5-1102M4 The true
 left end of clone RP11-13B20 is at 140238 in this sequence. The
 true right end of clone RP11-401M16 is at 25073 in this sequence.
 The true right end of clone RP11-72P17 is at 92766 in this
 sequence.

	Location/Qualifiers
1.	.150434
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP5-1102M4"
	/clone_lib="RPI-5"
t_region	1285..1588
	/note="AluX repeat: matches 1..304 of consensus"
t_region	3617..3928
	/note="Alu repeat: matches 1..311 of consensus"
t_region	6667..6964
	/note="AluX repeat: matches 1..299 of consensus"
t_region	6981..7111
	/note="FIAM_C repeat: matches 1..133 of consensus"
t_region	7445..7492
	/note="L1M3e repeat: matches -78..-28 of consensus"
t_region	8876..9427
	/note="L2 repeat: matches 1837..2451 of consensus"
t_region	9685..9886
	/note="MER20 repeat: matches 1..217 of consensus"
t_region	10646..110959
	/note="AluJo repeat: matches 1..312 of consensus"
t_region	11591..11731
	/note="AluJb repeat: matches 163..304 of consensus"
feature	12825..13546
	/note="Cpg Island"
	/evidence=not_experimental
t_region	14747..15040
	/note="AluX repeat: matches 1..296 of consensus"
t_region	15236..15371
	/note="MIR repeat: matches 86..242 of consensus"
t_region	16815..16935
	/note="MIR repeat: matches 23..150 of consensus"
t_region	17021..17192
	/note="AluSrx repeat: matches 132..303 of consensus"
t_region	17193..17500
	/note="AluX repeat: matches 1..307 of consensus"
t_region	17850..17923

JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970368.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA72P17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 154320 bases at least Q40
Consensus quality: 154754 bases at least Q20
Consensus quality: 154919 bases at least Q20
Insert size: 155332; sum-of-contigs
Insert size: 154076; 11.6% error; agarose-fp
Quality coverage: 12.87x in Q20 bases; sum-of-contigs Quality
coverage: 13.38x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source 1..155332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-72P17"
/clone_lib="RPC1-11.1"
misc_feature 1..155332
/note="assembly_fragment:02704
clone_end:T7
vector_side:left"
BASE COUNT 39582 a 39644 c 39869 g 36237 t
ORIGIN

Query Match 68.0%; Score 23.8; DB 2; Length 155332;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ccgccttcaatctgctcattaccaccccc 35
||||| ||||||| ||||||| |||||||
Db 117555 CCGCCTCTCATCTGTCATCCATTATCCCCCCCAC 117521

Search completed: February 20, 2002, 14:25:50
Job time: 14389 sec

Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: d1110L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Consensus quality: 151014 bases at least Q40
Consensus quality: 151075 bases at least Q30
Consensus quality: 151114 bases at least Q20
Insert size: 151149; sum-of-contigs
Insert size: 125492; 4.4% error; agarose-fp
Quality coverage: 10.38x in Q20 bases; sum-of-contigs Quality
coverage: 12.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 22192: contig of 22192 bp in length
* 22193 22292: gap of 100 bp
* 22293 68699: contig of 46407 bp in length
* 68700 68799: gap of 100 bp
* 68800 151349: contig of 82550 bp in length.

FEATURES
source 1..151349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q41-42.3"
/clone="RP5-1110L4"
/clone_lib="RPC1-5"
misc_feature 1..22192
/note="assembly_fragment:00611
fragment_chain:1"
22293..68699
/note="assembly_fragment:02237
fragment_chain:1"
68800..151349
/note="assembly_fragment:01926"
BASE COUNT 40982 a 37212 c 36594 g 36359 t 202 others
ORIGIN

Query Match 68.0%; Score 23.8; DB 2; Length 151349;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ccgccttcaatctgctcattaccaccccc 35
||||| ||||||| ||||||| |||||||
Db 124801 CCGCCTCTCATCTGTCATCCATTATCCCCCCCAC 124767

RESULT 15
AL365439/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-72P17, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL365439
VERSION AL365439.9 GI:15020472
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155332)
AUTHORS Martin S.
TITLE Direct Submission

us-09-904-420a-1_copy_851_885.rge

Thu Feb 21 07:25:39 2002

Thu Feb 21 07:25:40 2002

us-09-904-420a-1_copy_851_885.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 14:30:58 ; Search time 366.07 Seconds
(without alignments)
81.969 Million cell updates/sec

Title: US-09-904-420A-1_COPY_851_885
Perfect score: 35
Sequence: 1 ccgccttcaatctgctcattaccaccccccc 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	57.1	2797	22 AAC84603	Human ZF3 protein
C 2	19.2	54.9	418	22 AAF64276	Novel human polyn
C 3	19	54.3	525	22 AA117684	Probe #7617 for ge
C 4	19	54.3	525	22 AA142618	Probe #11304 used
C 5	19	54.3	650	22 AAH22402	Rat RISKMARKER2 co
6	19	54.3	1041	18 AAH77781	Nuclear mitotic ap
7	19	54.3	1092	18 AAH78310	EGAP-binding prote
8	19	54.3	1774	22 AAH99647	Human protein enco
9	19	54.3	1786	21 AAH21744	Human breast and o
10	19	54.3	2337	21 AAH39581	Human CRPL nucleot
11	19	54.3	2457	22 AAH58345	Human polynucleoti

12	19	54.3	2548	22	AAI60131	Human polynucleoti
13	19	54.3	4984	21	AAO7842	Nucleotide sequenc
14	19	54.3	6004	21	AAO7835	Human homologue of
C 15	18.8	53.7	91	17	AAI14299	IgE binding ligand
C 16	18.8	53.7	91	17	AAI14301	IgE binding ligand
C 17	18.8	53.7	337	21	AAH1961	N. meningitidis pa
C 18	18.8	53.7	487	22	AAI53718	Probe #22404 used
C 19	18.8	53.7	537	22	AAH09455	Human cDNA clone (
C 20	18.8	53.7	595	22	AAI140642	Probe #9328 used t
21	18.8	53.7	2122	22	AAH17156	Human cDNA sequenc
22	18.8	53.7	2405	22	AAH02351	Human RNA metaboli
C 23	18.8	53.7	3906	22	AAH67081	C glutamicum codin
C 24	18.8	53.7	349980	22	AAH68530	C glutamicum codin
C 25	18.6	53.1	239	22	AAH27148	Human leptin gene
C 26	18.6	53.1	319	14	AAQ60341	Human brain expres
C 27	18.6	53.1	362	16	AAH25059	Human gene signatu
C 28	18.6	53.1	485	21	AAC28272	Human secreted pro
C 29	18.6	53.1	654	21	AAH12845	Aspergillus oryzae
C 30	18.6	53.1	710	20	AAH21230	polynucleotide seq
C 31	18.6	53.1	999	22	AAH10144	Mouse lysosomal th
C 32	18.6	53.1	1187	17	AAT39536	E. coli lacI methy
C 33	18.6	53.1	1188	17	AAT39535	E. coli lacI methy
C 34	18.6	53.1	3426	20	AAH59851	SEQ ID 4 of W09925
C 35	18.6	53.1	3552	20	AAH32835	Sequences of H. fe
C 36	18.6	53.1	6830	18	AAT89255	Mouse 22B/30B (can
C 37	18.6	53.1	11817	18	AAT74199	Murine lystJ long
C 38	18.6	53.1	534720	19	AAV30458	Rhizobium species
C 39	18.6	53.1	536165	19	AAV30459	Rhizobium species
C 40	18.4	52.6	333	21	AAH02284	Human secreted pro
C 41	18.4	52.6	3973	20	AAH65326	Human neuronatin g
C 42	18.2	52.0	545	22	AAH13490	Human cDNA clone (
C 43	18.2	52.0	588	22	AAH39655	Probe #8341 used t
C 44	18.2	52.0	597	21	AAH75035	Human ORFX ORF590
C 45	18.2	52.0	2498	22	AAH18310	Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAC84603/c AAC84603 standard; DNA; 2797 BP.
ID AAC84603;
AC AAC84603;
DT 02-APR-2001 (first entry)
XX Human ZF3 protein encoding DNA.
DE
XX S-phase kinase associated protein; SKP1; SKP2; SKP2-like protein; ZF;
KW CUL-1; cullin; CDC33; p27; cyclin B; Max; Mad; c-Myc; MDM2; p53; Bax;
KW Bad; Bcl-2; tumour; cytosstatic; ds.
XX
OS Homo sapiens.
XX
XX WO2000075184-A1.
XX
XX 14-DEC-2000.
XX
XX 05-JUN-2000; 2000WO-US15449.
PF
PR 04-JUN-1999; 99US-0137494.
XX
XX (UYA) UNIV YALE.
XX
XX Zhang H, Tsvetkov LM, Kondo T;
XX
XX WPI; 2001-061703/07.
XX
XX P-PSDB; AAB48291.
XX
XX Modulating polypeptide levels in a cell, diagnosing and treating tumor,
XX
XX involves altering levels of proteins such as S-phase kinase associated
XX
XX proteins 1, 2 and cullin/CDC53 proteins .

Thu Feb 21 07:25:40 2002

XX Examples: Page 115-116; 162pp; English.
 XX The invention relates to methods of altering the polypeptide levels in a
 CC cell, using proteins selected from S-phase kinase associated proteins 1
 CC and 2 (SKP1, SKP2), SKP2-like proteins (ZF) and CUL-1 (a member of the
 CC cullin/ CUL5 family of proteins). The method is useful for altering the
 CC level of p27, cyclin E, Max, Mad, c-Myc, MDM2, p53, Bax, Bad or Bcl-2
 CC polypeptide in a cell. SKP2 and SKP2-like protein levels are useful for
 CC detecting tumours, and in monitoring tumor treatment in a mammal. Agents
 CC that modulate interactions between SKP and target proteins are useful for
 CC treating tumours.
 XX Sequence 2797 BP; 807 A; 525 C; 582 G; 883 T; 0 other;

Query Match 57.1%; Score 20; DB 22; Length 2797;
 Best Local Similarity 82.1%; Pred. No. 52;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 ttaaatctgtcattcattaccacccc 33
 DB 183 TTCATCTGCTCATGCTTTTGAACTCC 156

RESULT 2
 ID AAF64276/c
 XX AAF64276;
 XX 09-APR-2001 (first entry)
 XX Novel human polynucleotide, SEQ ID NO: 32.
 DE Human: cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.
 KW Homo sapiens.
 OS WO200102568-A2.
 XX 11-JAN-2001.
 XX 30-JUN-2000: 2000WO-US18374.
 XX 02-JUL-1999; 99US-0142310.
 XX 02-JUL-1999; 99US-0142311.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Ieshkowitz D;
 PI Kita D, Garcia V, Jones LW, Strache-Crain B;
 XX WPI: 2001-091805/10.
 XX Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 XX Claim 9; Page 549; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and

CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

XX Sequence 418 BP; 89 A; 93 C; 117 G; 114 T; 5 other;

Query Match 54.9%; Score 19.2; DB 22; Length 418;
 Best Local Similarity 75.0%; Pred. No. 79;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 ccttcattctgtcattcattaccacccc 35
 DB 218 COTTCATCTCTCTAATCCACCCCTTACCCCC 187

RESULT 3
 ID AAI17684/c
 XX AAI17684;
 XX 12-OCT-2001 (first entry)
 XX Probe #7617 for gene expression analysis in human cervical cell sample.
 DE Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 KW Homo sapiens.
 OS WO200157278-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001: 2001WO-US00670.
 XX 04-FEB-2000: 2000US-0180312.
 XX 26-MAY-2000: 2000US-0207456.
 XX 30-JUN-2000: 2000US-0608408.
 XX 03-AUG-2000: 2000US-0632366.
 XX 21-SEP-2000: 2000US-0234687.
 XX 27-SEP-2000: 2000US-0236359.
 XX 04-OCT-2000: 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI: 2001-488901/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX Claim 25; SEQ ID No 7617; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 525 BP; 117 A; 134 C; 82 G; 192 T; 0 other;

```

Query Match      54.3%; Score 19; DB 22; Length 525;
Best Local Similarity 81.5%; Pred. No. 99;
Matches 22; Conservative 0; Mismatches 0; Gaps 0;

Yy 1 ccgccttcaatcgtcatgcatcattacc 27
    I | | | | | | | | | | | | | | | |
Db 513 CTGCCTTCAATCGCTTGATGATCTACC 487

RESULT 4
AAH22402 ID AA142618 standard; DNA; 525 BP.
XX AC
XX AAC
XX TTT
XX TTT
DE 17-OCT-2001 (first entry)
XX Probe #11304 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KK genetic disorder; ss.
XX Homo sapiens.
XX OS
XX PN
XX PN
XX XX
XX PD
XX PF
XX PF
XX 30-JAN-2001; 2001WO-US006663.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
DR analyzing gene expression in human placenta .
XX PT
XX Claim 25; SEQ ID NO 11304; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX XX
XX Sequence 525 BP; 117 A; 134 C; 82 G; 192 T; 0 other;
SQ

Query Match      54.3%; Score 19; DB 22; Length 525;
Best Local Similarity 81.5%; Pred. No. 99;
Matches 22; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 ccgccttcaatcgtcatgcatcattacc 27
    I | | | | | | | | | | | | | | | |
Db 513 CTGCCTTCAATCGCTTGATGATCTACC 487

RESULT 5
AAH22402 ID AAH22402 standard; DNA; 650 BP.
XX XX
XX AAH22402.
```

Thu Feb 21 07:25:40 2002

```

KW malignant cell growth; ss.
XX
OS Homo sapiens.
XX
PN WO9640917-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09504.
XX
PR 07-JUN-1995; 95US-0478408.
XX
PA (UYUA ) UNIV YALE.
XX
PI McPherson SMG, Snyder MP;
XX
PT WPI; 1997-077270/07.
XX
DR P-PSDB; AAW21730.
XX
PT New nucleic acid encoding nuclear mitotic appts. interacting
PT proteins - useful for modulating cell division and proliferation and
PT in diagnosis
XX
PS Claim 2; Page 64-65; 78pp; English.
XX
CC The sequences given in AAT7780-81 encode NIP-1 and NIP-2 (NIP = Numa
CC (nuclear mitotic apparatus) interacting protein). Compounds which
CC interfere with the interaction of Numa with a known NIP are used to
CC modulate cell division and/or proliferation. Ab. raised conventionally
CC using NIP-1 or -2 as immunogen, are used to detect NIP (or their
CC complexes) and to block their activity for diagnostic or therapeutic
CC use, e.g. to detect defective Numa or NIP which may be markers for
CC aberrant (including malignant) cell growth (which can also be detected
CC by nucleic acid sequencing). Also where malignancy is related to defects
CC in Numa or NIP, it can be treated by admin. of the appropriate
CC functional protein.
XX
SQ Sequence 1041 BP; 245 A; 311 C; 314 G; 171 T; 0 other;

Query Match 54.3%; Score 19; DB 18; Length 1041;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ccgccttcaatctgctcattaccaccccc 35
DB 259 ccgcctgcccccaacctcaggcatgtcccatcccc 293

RESULT 7
AAT78310
XX AAT78310 standard; cDNA; 1092 BP.
XX
AC AAT78310;
XX
DT 10-OCT-1997 (first entry)
XX
DE E6AP-binding protein cln42 encoding cDNA.
XX
KW Human; cell differentiation; survival; carcinoma; psoriasis;
KW ubiquitination; p53; tumour suppressor; homeostasis; papilloma virus;
KW epithelial cell; acne; ichthyosis; aphthous ulcer; hair growth;
KW antibody; cell proliferation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1092
FT FT /*tag= a
FT FT /product= cln42
FT FT /note= "E6AP-binding protein"
XX
PN WO9640767-A2.

```

```

XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09040.
XX
XX 07-JUN-1995; 95US-0484878.
XX
XX (MITO-) MITOTIX INC.
XX
XX Beer-Romero PL, Draetta G, Rolfe M;
XX
XX WPI; 1997-087053/08.
XX
XX P-PSDB; AAW23658.
XX
XX E6AP-binding proteins and related nucleic acid - useful for
XX modulating cell differentiation, survival etc., partic. for
XX treatment and diagnosis of carcinoma, psoriasis, etc
XX
XX Claim 12; Page 66-68; 83pp; English.
XX
CC The present sequence encodes the E6AP-binding protein cln42. E6AP
CC mediates ubiquitination and so the inactivation of e.g. p53 tumour
CC suppressor. The E6AP-binding protein is likely to be involved in normal
CC cell homeostasis and in the pathogenesis of proliferative and
CC differentiation disorders, e.g. regulation of gene expression or the
CC cell cycle, modification of cell surface receptors, biogenesis of
CC ribosomes and DNA repair. The protein, which can optionally be generated
CC in vivo by gene therapy, may be useful in treatment and prevention of
CC papilloma virus infected epithelial cell processes more generally, e.g. in
CC cases of psoriasis, acne, ichthyosis, aphthous ulcers. It may also be
CC used to control wound healing, and inhibit growth of hair. It can also
CC be used to generate antibodies which are used in immunoassays to
CC determine the protein levels. The encoding nucleic acid can be used to
CC prepare recombinant proteins and oligonucleotides useful as probes and
CC primers for diagnostic detection of (mutant) mRNA for the protein in
CC (transformed) cells, also for antisense therapy. It can also be used
CC for the detection of mutations in E6AP-binding protein-encoding genes,
CC mis-expression of these genes or mis-incorporation of the protein in a
CC transcription regulatory complex containing E6AP can be used to assess
CC risk of disorders characterised by cell proliferation.
XX
XX Sequence 1092 BP; 260 A; 329 C; 324 G; 179 T; 0 other;

Query Match 54.3%; Score 19; DB 18; Length 1092;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ccgccttcaatctgctcattaccaccccc 35
DB 310 ccgcctgcccccaacctcaggcatgtcccatcccc 344

RESULT 8
AAH99647
XX ID AAH99647 standard; cDNA; 1774 BP.
XX
XX AAH99647;
XX
XX 16-OCT-2001 (first entry)
XX
XX Human protein encoding cDNA sequence SEQ ID NO:482.
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiac; central nervous system; virucide;
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX antiaggregant; haemostatic; vulnery; antidiabetic; eczema;
XX dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

```


Thu Feb 21 07:25:40 2002

us-09-904-420a-1_copy_851_885.rng

DT 28-FEB-2000 (first entry)
 DE Human cREL nucleotide sequence.
 XX
 KW Human: cREL; transcriptional activator; antisense compound;
 KW therapeutic; ss.
 KW
 OS Homo sapiens.
 XX
 PN US6001652-A.
 XX
 PD 14-DEC-1999.
 XX
 PF 18-SEP-1998; 98US-0156253.
 XX
 PR 18-SEP-1998; 98US-0156253.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Monia BP, Cowser LM, Baker BF;
 DR WPI; 2000-061889/05.
 DR P-PSDB; AAY55804.
 XX
 PT Antisense modulation of human cREL expression -
 PS Example 13; Column 31; 26pp; English.
 XX
 CC The invention provides antisense compounds targeted to a coding region,
 CC 3'-UTR or 5'-UTR of a nucleic acid molecule encoding human cREL
 CC (transcriptional activator). The antisense compounds are useful as
 CC research agents and diagnostics such as in the elucidation of the
 CC function of a particular gene. The antisense compounds can be useful as
 CC therapeutic modalities that can be configured to be useful in treatment
 CC regimes for treatment of cells, tissues and animals, especially humans.
 CC In the prior art, there are no known therapeutic agents which effectively
 CC inhibit the synthesis of cREL and additional agents capable of inhibiting
 CC cREL function are still required. The present sequence represent the
 CC human cREL sequence (GenBank Accn No: X75042).
 XX
 SQ Sequence 2337 BP; 722 A; 491 C; 524 G; 598 T; 2 other;
 XX
 Query Match 54.3%; Score 19; DB 21; Length 2337;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 cgccttcgaatgtgctgcatgattaccaccccc 35
 Db 1399 cctcttcaagctggtcctcagtgagccaccccc 1433
 RESULT 11
 AAI58345
 ID AAI58345 standard; cDNA; 2457 BP.
 XX
 AC AAI58345;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 548.
 XX
 KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX

PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM39189.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -
 PS Claim 1; SEQ ID NO 548; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAAM42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2457 BP; 768 A; 517 C; 555 G; 617 T; 0 other;
 XX
 Query Match 54.3%; Score 19; DB 22; Length 2457;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 cgccttcgaatgtgctgcatgattaccaccccc 35
 Db 1300 cctcttcaagctggtcctcagtgagccaccccc 1334
 RESULT 12
 AAI60131
 ID AAI60131 standard; cDNA; 2548 BP.
 XX
 AC AAI60131;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 4120.
 XX
 KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX

[illegible]

Thu Feb 21 07:25:40 2002

FT /*tag= a
FT /note= "this region is found to be absent in a
FT variant cDNA isolated from frontal cortex"
FT 3098..3121
FT /*tag= b
FT /note= "this region is found to be absent in a
FT variant cDNA isolated from frontal cortex"
FT 3518..3526
FT /*tag= c
FT /note= "this region is found to be absent in a
FT variant cDNA from Hela or colorectal
FT adenocarcinoma tissue"
XX
XX WO9963080-A1.
XX
XX 09-DEC-1999.
XX
XX 02-JUN-1999; 99WO-EP03848.
XX
XX 03-JUN-1998; 98GB-0011962.
XX
XX (JANC) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
XX Maerten LJS, Verhasselt P, Van De Craen M;
XX
XX WPI; 2000-116370/10.
XX
XX P-PSDB; AAY85564.
XX
XX Novel proteins and nucleic acids e.g. for treating neurodegeneration -
XX
XX Claim 92; Fig 1a; 146pp; English.
XX
XX The invention provides vertebrate (human) protein homologue of a UNC-53
XX protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
XX their plus ends. The UNC-53 sequences are used to promote neural
XX regeneration, revascularization and wound healing; also for treating
XX neurodegenerative diseases, acute traumatic injury, fibrotic disease and
XX autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
XX polynucleotides can be used for recombinant production of the proteins,
XX as a source of probes for detecting allelic variants and polymorphisms,
XX for sequencing genomic DNA and for detecting UNC-53 expression; and as
XX source of therapeutic antisense sequences. Cells that express the
XX protein are used to identify regulators of cell shape, growth, motility
XX and migration. They can also be used to identify proteins that are
XX involved in signal transduction pathways also involving UNC-53, and to
XX identify compounds that alter attachment of UNC-53 to microtubules. A
XX target gene coupled to a UNC-53 encoding sequence may be used to deliver
XX the target gene to a cellular microtubule or its plus ends. The present
XX sequence represents the nucleotide sequence of the first human homologue
XX of UNC-53, designated Hs-UNC-53/1.
XX
XX Sequence 6004 BP; 1470 A; 1844 C; 1562 G; 1127 T; 1 other;

Query Match 54.3%; Score 19; DB 21; Length 6004;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 cgcgccttcattgctgcattaccaccacccccc 35
||||| ||| |||| || ||| |||
Db 120 cgcgccttcattgctgcattaccaccacccccc 154

RESULT 15
AAT14299/c
XX AAT14299 standard; RNA; 91 BP.
XX AAT14299;
XX
XX 21-NOV-1996 (first entry)
XX
XX IgE binding ligand IGEL 47.1.

XX Immunoglobulin E; RNA ligand; inhibitor; IgE receptor; therapy; asthma;
KW IgE dependent reaction; allergic disease; allergic rhinitis; hay fever;
KW atopic dermatitis; chronic skin irritation; anaphylactic shock; IgE; ss.
XX Synthetic.
OS
XX Key Location/Qualifiers
FH modified_base 1..91
FT /*tag= a
FT /note= "all C's are 2'NH2 cytosine, and all U's are 2'NH2
FT uracil"
XX
XX WO9610576-A1.
XX
XX 11-APR-1996.
XX
XX 27-SEP-1995; 95WO-US12401.
XX
XX 06-JUN-1995; 95US-0471985.
XX 03-OCT-1994; 94US-0317403.
XX
XX (NEXS-) NEXSTAR PHARM INC.
XX
XX Gold L, Tasset D, Wiegand TW;
XX WPI; 1996-209316/21.
XX
XX Identification of high-affinity IgE oligo:nucleotide ligands - which
XX inhibit interaction of IgE with its receptor, which are therefore
XX useful in diagnostic and therapeutic applications
XX
XX Claim 9; Page 39; 88pp; English.
XX
XX AAT14280-T14320 represent RNA ligands for immunoglobulin E (IgE). These
XX sequences are non-naturally occurring RNA ligands, and were isolated and
XX purified using the method of the invention. In this method, a candidate
XX mixture of nucleic acids is prepared, and contacted with IgE. The
XX sequences with an increased affinity to IgE relative to the mixture are
XX partitioned off from the rest of the mixture. The partitioned sequences
XX are then amplified, to yield a mixture of nucleic acids enriched for
XX those with relatively higher affinity and specificity for IgE binding.
XX The steps of this method can be repeated in order to obtain a mixture of
XX higher specificity for IgE binding. This method can be carried out with
XX RNA or DNA. AAT14240-T14243, and AAT14431-T14487 represent the DNA
XX ligands identified by this method. The IgE ligands inhibit the function
XX of IgE by preventing the IgE/receptor interaction. The ligands are
XX therefore useful in diagnostic and therapeutic applications for diseases
XX associated with IgE dependent reactions. The IgE dependent reactions
XX include allergic diseases such as allergic rhinitis (hay fever), asthma,
XX atopic dermatitis (chronic skin irritations), and anaphylactic shock.
XX
XX Sequence 91 BP; 22 A; 19 C; 35 G; 15 U; 0 other;

Query Match 53.7%; Score 18.8; DB 17; Length 91;
Best Local Similarity 76.7%; Pred. No. 88;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 ttcaatctgctgcattaccaccacccccc 35
||||| ||| |||| || ||| |||
Db 47 TTCACCTCTGTCACCTCATTATCCGCCCCAC 18

Search completed: February 20, 2002, 14:31:01
Job time: 12450 sec

us-09-904-420a-1_copy_851_885.rng

Thu Feb 21 07:25:40 2002

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	54.3	2337	3	US-09-156-253-1	Sequence 1, Appli
C 2	18.8	53.7	91	1	US-08-317-403A-26	Sequence 26, Appli
C 3	18.8	53.7	91	1	US-08-317-403A-28	Sequence 28, Appli
C 4	18.8	53.7	91	1	US-08-471-985A-26	Sequence 26, Appli
C 5	18.8	53.7	91	1	US-08-471-985A-28	Sequence 28, Appli
C 6	18.8	53.7	91	5	PCR-US95-12401A-26	Sequence 26, Appli
C 7	18.8	53.7	91	5	PCR-US95-12401A-28	Sequence 28, Appli
C 8	18.6	53.1	239	4	US-09-437-457-17	Sequence 17, Appli
C 9	18.6	53.1	1187	5	PCR-US95-13356-2	Sequence 2, Appli
C 10	18.6	53.1	1188	5	PCR-US95-13356-1	Sequence 1, Appli
C 11	18.6	53.1	3552	4	US-09-157-210-3	Sequence 3, Appli
C 12	18.6	53.1	6830	2	US-08-822-445-1	Sequence 1, Appli
C 13	18.4	52.6	3973	2	US-08-602-093-6	Sequence 6, Appli
C 14	18	51.4	1854	3	US-08-923-454A-13	Sequence 13, Appli
C 15	18	51.4	2800	3	US-08-448-138-3	Sequence 3, Appli
C 16	18	51.4	3537	2	US-08-363-124A-3	Sequence 3, Appli
C 17	18	51.4	3652	3	US-08-936-135-5	Sequence 5, Appli
C 18	17.8	50.9	7091	4	US-08-975-762-46	Sequence 46, Appli
C 19	17.8	50.9	7091	4	US-09-295-028-46	Sequence 46, Appli
C 20	17.8	50.9	7091	4	US-09-106-582-46	Sequence 46, Appli
C 21	17.8	50.9	8342	5	US-08-545-860D-63	Sequence 63, Appli
C 22	17.8	50.9	8342	5	PCR-US94-04496-63	Sequence 63, Appli
C 23	17.8	50.9	8473	4	US-09-172-422-2	Sequence 2, Appli
C 24	17.6	50.3	1244	2	US-08-755-728-1	Sequence 1, Appli
C 25	17.6	50.3	1244	2	US-08-974-655-1	Sequence 1, Appli
C 26	17.6	50.3	1244	4	US-09-283-011-1	Sequence 1, Appli
C 27	17.6	50.3	1281	2	US-09-016-000-7	Sequence 7, Appli

RESULT 2
US-08-317-403A-26/c
; Sequence 26, Application US/08317403A
; Patent No. 5629155
; GENERAL INFORMATION:
; APPLICANT: WIEGAND, T., GOLD, L., AND TASSET, D.
; TITLE OF INVENTION: HIGH-AFFINITY
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS
; TITLE OF INVENTION: TO IMMUNOGLOBIN E
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite 200
; CITY: Englewood

Thu Feb 21 07:25:41 2002

```
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,403A
; FILING DATE: 3-OCT-1994
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX23
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
;
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
;
; US-08-317-403A-26
;
;
; Query Match 53.7%; Score 18.8; DB 1; Length 91;
; Best Local Similarity 76.7%; Pred. No. 19;
; Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
; QY 6 ttcaatctgctatgcattaccaccccc 35
; Db 47 TTCACCTGTCTACTCATTATCCGCCCCAC 18
;
; RESULT 3
; US-08-317-403A-28/C
; Sequence 28, Application US/08317403A
; Patent No. 5629155
;
; GENERAL INFORMATION:
; APPLICANT: WIEGAND, T., GOLD, L., AND TASSET, D.
; TITLE OF INVENTION: HIGH-AFFINITY
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS
; TITLE OF INVENTION: TO IMMUNOGLOBIN E
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,403A
; FILING DATE: 3-OCT-1994
; CLASSIFICATION: 435
```

```
;
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX23
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
;
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
;
; US-08-317-403A-28
;
;
; Query Match 53.7%; Score 18.8; DB 1; Length 91;
; Best Local Similarity 76.7%; Pred. No. 19;
; Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
; QY 6 ttcaatctgctatgcattaccaccccc 35
; Db 47 TTCACCTGTCTACTCATTATCCGCCCCAC 18
;
; RESULT 4
; US-08-471-985A-26/C
; Sequence 26, Application US/08471985A
; Patent No. 5686592
;
; GENERAL INFORMATION:
; APPLICANT: WIEGAND, Torsten Walter
; APPLICANT: GOLD, Larry
; APPLICANT: TASSET, Diane
; TITLE OF INVENTION: HIGH-AFFINITY
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS TO
; TITLE OF INVENTION: IMMUNOGLOBIN E
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,985A
; FILING DATE: 06-JUNE-1995
; CLASSIFICATION: 536
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/317,403
; FILING DATE: 03-OCTOBER-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
;
; ATTORNEY/AGENT INFORMATION:
```

NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX23/CIP
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 91
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 cytosine
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 uracil
US-08-471-985A-28

Query Match 53.7%; Score 18.8; DB 1; Length 91;
Best Local Similarity 76.7%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 ttcaatctgctcatgcattaccaccccc 35
||||| ||||| ||||| ||||| ||||| |||||
Db 47 TTCACCTCTGTTCACTCATTAATCCGCCCCAC 18

RESULT 5
PCT-US95-12401A-28/c
Sequence 28, Application US/08471985A
Patent No. 5686592
GENERAL INFORMATION:
APPLICANT: WIEGAND, Torsten Walter
APPLICANT: GOLD, Larry
APPLICANT: TASSET, Diane
TITLE OF INVENTION: HIGH-AFFINITY
TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS TO
TITLE OF INVENTION: IMMUNOGLOBIN E
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 East Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MG storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,985A
FILING DATE: 06-JUNE-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/317,403
FILING DATE: 03-OCTOBER-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX23/CIP
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433

NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX23/CIP
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 91
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 cytosine
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 uracil
US-08-471-985A-28

Query Match 53.7%; Score 18.8; DB 1; Length 91;
Best Local Similarity 76.7%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 ttcaatctgctcatgcattaccaccccc 35
||||| ||||| ||||| ||||| ||||| |||||
Db 47 TTCACCTCTGTTCACTCATTAATCCGCCCCAC 18

RESULT 6
PCT-US95-12401A-26/c
Sequence 26, Application PC/TUS9512401A
GENERAL INFORMATION:
APPLICANT: WIEGAND, TORSTEN
APPLICANT: GOLD, LARRY
APPLICANT: TASSET, DIANE
TITLE OF INVENTION: HIGH-AFFINITY
TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS
TITLE OF INVENTION: TO IMMUNOGLOBIN E
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 East Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MG
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12401A
FILING DATE: 27 SEPTEMBER 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/471,985
FILING DATE: 06-JUNE-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/317,403
FILING DATE: 03-OCTOBER-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX23C/PCT
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 91

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: RNA
: FEATURE:
: OTHER INFORMATION: All C's are 2'-NH2 cytosine
: FEATURE:
: OTHER INFORMATION: All U's are 2'-NH2 uracil
PCT-US95-12401A-26

```

Query Match 53.7%; Score 18.8; DB 5; Length 91;
Best Local Similarity 76.7%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 ttcaatctgctcatgcattaccaccccc 35
||||| ||| ||||| ||| ||||| |
pb 47 ttcaactctgttcaactcattatccgccccac 18

```

RESULT 7
PCT-US95-12401A-28/c
; Sequence 28, Application PC/TUS9512401A
; GENERAL INFORMATION:
;
; APPLICANT: WIEGAND, TORSTEN
; APPLICANT: GOLD, LARRY
; APPLICANT: TASSET, DIANE
; TITLE OF INVENTION: HIGH-AFFINITY
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIGANDS
; TITLE OF INVENTION: TO IMMUNOGLOBIN E
;
; NUMBER OF SEQUENCES: 112
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111

```

```

; MOLECULE TYPE: RNA
;
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
;
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
;
DCT-US95-12401A-28

```

Query Match 53.7%; Score 18.8; DB 5; Length 91;
Best Local Similarity 76.7%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 6 ttcaatctgctcatgcaltacccaccccc 35
||||| ||||| ||||| ||||| ||||| ||
nb 47 TTCACCTGTTCACCTCATTTATCCGCCCCC 18

```

RESULT      8
US-09-437-457-17/c
; Sequence 17, Application US/09437457
; Patent No. 6273893
; GENERAL INFORMATION:
; APPLICANT: Giordano, Anthony
; APPLICANT: Xavier, Ashish
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES AND METHODS FOR
; TITLE OF INVENTION: IDENTIFYING COMPOUNDS THAT AFFECT RNA/RNA BINDING PROTEIN
; TITLE OF INVENTION: IDENTIFYING COMPOUNDS THAT AFFECT RNA/RNA BINDING PROTEIN
; TITLE OF INVENTION: INTERACTIONS AND MRNA FUNCTIONALITY
; FILE REFERENCE: 50093/014001
; CURRENT APPLICATION NUMBER: US/09/437,457
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-437-457-17

```

Query Match 53.1%; Score 18.6; DB 4; Length 239;
Best Local Similarity 72.7%; Pred. No. 28;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ccgccttcaatctgctaatgcattaccacccc 33
|| || | |||| | ||| |||||
pb 167 cctcttccctcttgcgcaaacattccccacccc 135

```

1  RESULT          9
2  PCT-US95-13536-2
3  ; Sequence 2, Application PC/TUS9513536
4  ; GENERAL INFORMATION:
5  ; APPLICANT: SKOPEK, THOMAS R
6  ; TITLE OF INVENTION: SYNTHESIS OF METHYLASE RESISTANT GENES
7  ; NUMBER OF SEQUENCES: 2
8  ; CORRESPONDENCE ADDRESS:
9  ; ADDRESSEE: ELLIOTT KORSEN
10 ; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
11 ; CITY: Rahway
12 ; STATE: NJ
13 ; COUNTRY: US
14 ; ZIP: 07065
15 ; COMPUTER READABLE FORM:
16 ; MEDIUM TYPE: Floppy disk
17 ; COMPUTER: IBM PC compatible
18 ; OPERATING SYSTEM: PC-DOS/MS-DOS
19 ; SOFTWARE: Patent In Release #1.0, Version #1.30
20 ; CURRENT APPLICATION DATA:
21 ; APPLICATION NUMBER: PCT/US95/13536
22 ; FILING DATE: 13-OCT-1995
23 ; CLASSIFICATION:
24 ; ATTORNEY/AGENT INFORMATION:
25 ; NAME: KORSEN, ELLIOTT

```

```
Query Match      53.1%  Score 18.6;  DB 5;  Length 1188;
Best Local Similarity 72.7%  Pred. No. 39;
Matches 24;  Conservative 0;  Mismatches 9;  Indels 0;  Gaps 0;
```

RESULT 12
US-08-822-445-1
: Sequence 1, Application US/08822445
: Patent No. 5952223
: GENERAL INFORMATION:
: APPLICANT: Kaplan, Jerry
: APPLICANT: Perou, Charles
: APPLICANT: Moore, Karen
: TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036/2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/822.445
: FILING DATE: 21-MAR-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30.742

us-09-904-420a-l_copy_85l_885.rni

Thu Feb 21 07:25:41 2002

REFERENCE/DOCKET NUMBER: 7853-062-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6830 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...6558
OTHER INFORMATION:
US-08-822-445-1

Query Match 53.1%; Score 18.6; DB 2; Length 6830;
Best Local Similarity 84.0%; Pred. No. 56;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 ccttcaatctgctcattacc 28
Db 4622 CTTCAATGATCTCATGACGACCC 4646

RESULT 13
US-08-602-093-6/c
; Sequence 6, Application US/08602093
; Patent No. 5837535
; GENERAL INFORMATION:
; APPLICANT: Joseph, Rajiv
; APPLICANT: Dou, Dexian
; TITLE OF INVENTION: A NOVEL NEURONAL-NEONATAL GENE:
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5837535thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,093
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 1059.00015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-602-093-6

Query Match 52.6%; Score 18.4; DB 2; Length 3973;
Best Local Similarity 78.6%; Pred. No. 61;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 tcaatctgctcattaccacccc 34
Db 1224 TCCATCTGCTCTTCCATACCGACCC 1197

RESULT 14
US-08-923-454A-13
; Sequence 13, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-923-454A-13

Query Match 51.4%; Score 18; DB 3; Length 1854;
Best Local Similarity 70.6%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 ccgccttcaatctgctcattaccacccc 34
Db 1596 CGCCATCTACCTGTTAATGACGTAATCTCCCTC 1629

RESULT 15
US-08-448-194-3/c
; Sequence 3, Application US/08448194

Thu Feb 21 07:25:41 2002

Patent No. 6028049
GENERAL INFORMATION:
APPLICANT: JACOBS, Eric
APPLICANT: LEGRAIN, Michele
APPLICANT: MAZARIN, Veronique
APPLICANT: BOUCHON-THEISEN, Bernadette
APPLICANT: SCHRYVERS, Anthony B.
APPLICANT: BLOCH, Marie-Aline
TITLE OF INVENTION: DNA FRAGMENTS CODING FOR THE TRANSFERRIN
RECEPTOR OF NEISSERIA MENINGITIDIS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,194
FILING DATE: 23-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,469
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/078,053
FILING DATE: 18-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92 07493
FILING DATE: 19-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 016100-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: DNA encodes Tpb1 subunit of transferrin
ORGANISM: receptor
STRAIN: Neisseria meningitidis IM2394
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 40..111
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 112..2763
FEATURE:
NAME/KEY: CDS
LOCATION: 40..2763
US-08-448-194-3

Search completed: February 20, 2002, 14:32:31
Job time: 14410 sec

Query Match 51.4%; Score 18; DB 3; Length 2800;
Best Local Similarity 70.6%; Pred No. 83;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 1 ccgccttcaatcgtcattaccaccccc 34
Db 749 CCGCATCTTTATGCGCATGGATTTCGCCGCC 716

Thu Feb 21 07:25:43 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2002, 13:52:30 ; Search time 3793.25 Seconds
(without alignments)
99.150 Million cell updates/sec

Title: US-09-904-420A-1_COPY_851_885
Perfect score: 35
Sequence: 1 ccgcctcaatctgcattgcattaccaccccc 35

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estfun.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estom.*
- 5: em_estpl.*
- 6: em_estba.*
- 7: em_estro.*
- 8: em_estov.*
- 9: em_htc.*
- 10: gb_estl.*
- 11: gb_est2.*
- 12: gb_htc.*
- 13: gb_gss.*
- 14: em_gss_fun.*
- 15: em_gss_hum.*
- 16: em_gss_inv.*
- 17: em_gss_pln.*
- 18: em_gss_pro.*
- 19: em_gss_rnd.*
- 20: em_gss_vrt.*
- 21: em_gss_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	201	11	BF799331 RC6-C1012
2	35	100.0	353	11	BG386035 60245204
3	33	94.3	273	10	AA311787 EST182503
4	27	77.1	264	10	BB584341 BB584341
5	27	77.1	400	11	BG555601 ia84a02.x
6	27	77.1	649	11	BI154476 602904810
7	26.2	74.9	288	10	A1049127 ub35d04.f
8	26.2	74.9	330	10	AA958693 v294h1.f
9	26.2	74.9	587	11	BF538482 602053985
10	22.4	64.0	516	10	AV387748 AV387748
11	22.2	63.4	465	10	A1234477 EST380 Ma
12	21.8	62.3	701	11	BF279037 GA_EB003

C	13	21.4	61.1	558	10	AU031552
C	14	21.4	61.1	612	11	BF521414
C	15	21.4	61.1	970	13	CNS003NE
C	16	21.2	60.6	298	13	AZ578207
C	17	21.2	60.6	470	13	AQ121606
C	18	21.2	60.6	555	10	AW974385
C	19	21.2	60.6	1010	10	BE284178
C	20	21	60.0	669	13	BE422877
C	21	21	60.0	1078	10	BE743309
C	22	20.8	59.4	152	10	AI1131784
C	23	20.8	59.4	244	10	BB033197
C	24	20.8	59.4	638	10	BE422877
C	25	20.8	59.4	735	10	AW181048
C	26	20.8	59.4	818	10	BE040637
C	27	20.8	59.4	861	11	BF026907
C	28	20.8	59.4	1441	12	AK011834
C	29	20.6	58.9	474	13	AZ049530
C	30	20.6	58.9	539	11	BE264604
C	31	20.6	58.9	632	11	RG167694
C	32	20.6	58.9	649	13	AZ053748
C	33	20.6	58.9	765	11	BG826357
C	34	20.6	58.9	797	11	BG828123
C	35	20.6	58.9	823	13	BE047537
C	36	20.6	58.9	866	11	BG397266
C	37	20.6	58.9	1014	11	BG330286
C	38	20.4	58.3	324	13	B17386
C	39	20.4	58.3	590	13	AZ452837
C	40	20.4	58.3	904	13	CNS040XE
C	41	20.4	58.3	1002	10	BF561073
C	42	20.2	57.7	253	10	AV331663
C	43	20.2	57.7	358	11	BF202406
C	44	20.2	57.7	361	13	AZ758698
C	45	20.2	57.7	520	11	BF279061

ALIGNMENTS

RESULT 1

LOCUS	BF799331	201 bp	mRNA	EST	12-JAN-2001
DEFINITION	RC6-C10122-091000-021-D10 C10122 Homo sapiens cDNA, mRNA sequence.				
ACCESSION	BF799331				
VERSION	BF799331.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 201) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai M.A., da Silva,W. Jr., Zaqui,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.C. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6612-RC6-C10122-091000-021-D10&t3=2000-10-09&t4=1) Seq primer: puc 18 forward				

Thu Feb 21 07:25:43 2002

High quality sequence start: 16
High quality sequence stop: 201.
Location/Qualifiers
1. .201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0122"
/dev_stage="Adult"
/note="Organ: colon;ins: Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
45 a 71 c 50 g 35 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 35; DB 11; Length 201;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgcctcaatctgcctcattaccaccccc 35
|||||
Db 122 CCGCCTTCATCTGCTCATGCATTACCCACCCCC 156

RESULT 2
LOCUS BG386035 353 bp mRNA EST 12-MAR-2001
DEFINITION 602455204F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583473 5',
mRNA sequence.
ACCESSION BG386035
VERSION BG386035.1 GI:13279481
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 353)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLCMI307 row: h column: 02
High quality sequence stop: 276.
Location/Qualifiers
1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583473"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
80 a 111 c 122 g 40 t

BASE COUNT
ORIGIN

Query Match 94.3%; Score 33; DB 10; Length 273;
Best Local Similarity 94.3%; Pred. No. 0.052;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 100.0%; Score 35; DB 11; Length 353;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgcctcaatctgcctcattaccaccccc 35
|||||
Db 224 CCGCCTTCATCTGCTCATGCATTACCCACCCCC 258

RESULT 3
LOCUS AA311787 273 bp mRNA EST 19-APR-1997
DEFINITION EST182503 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
similar to T-cell factor 1, A/B/C, mRNA sequence.
ACCESSION AA311787
VERSION AA311787.1 GI:1964114
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 273)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-L., Macraes, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Otterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
, Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hung, J., Li, H., Meissner, P., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/ldb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1. .273
/organism="Homo sapiens"
/db_xref="ATCC (inhost):158998"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells VI"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
60 a 111 c 59 g 39 t 4 others

BASE COUNT
ORIGIN

Query Match 94.3%; Score 33; DB 10; Length 273;
Best Local Similarity 94.3%; Pred. No. 0.052;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

[illegible]

musculus cDNA clone 9230105F22 5', mRNA sequence.

SOURCE	ORGANISM
house mouse.	
Mus musculus	
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	

REFERENCE
AUTHORS
1 (bases 1 to 264)
Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodyama, Y.,

TITLE
Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
R1EN Mouse ESTs (Aizawa, K. et al. 2000)

Unpublished (2000)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
of the Institute of Physical and Chemical Research (RIKEN)

the Institute of Physical and Chemical Research (Kisarazu, Japan)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Kicsunai, F., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, F., Ozawa, Y., Muramatsu, M., Okazaki,

Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FFEATURES

source

1..264

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="9330105F22"

```

/clone_lib="RIKEN full-length enriched, adult male
epididymis"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Site 1: Sall; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'

```

oligo-dT priming. XhoI site destroyed during cloning.
Size-selected by column fractionation; average insert size
1.2kb. Primary library, unamplified."

BASE COUNT ORIGIN

87 a 148 c 98 g 67 t

Query Match 77.1%; Score 27; DB 11; Length 400;

Best Local Similarity 85.7%; Pred. No. 6.9;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgccttcaatctgctcatgcatcaccaccccc 35

|||||

Db 138 CTGCTTCAATCTGCTCATGCCCTACCCACCGGCC 172

RESULT 6

BI154476 649 bp mRNA EST 05-JUL-2001
LOCUS 602904810F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5034194 5',
mRNA sequence.

ACCESSION BI154476

VERSION BI154476.1 GI:14614477

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 649)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:

<http://image.llnl.gov>

Plate: LLAM1095 row: d column: 03

High quality sequence stop: 647.

FEATURES

source

1..649

Location/Qualifiers

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:5034194"

/clone_lib="NIH_CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert 2 kb. Library constructed by Life

Technologies, catalog #12017-018. Investigators providing

samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference

for transgenic model: Xu et al., Nature Genetics 22, 37-43

(1999). Note: this is a NCI_CGAP Library."

BASE COUNT

ORIGIN

Query Match 77.1%; Score 27; DB 11; Length 649;

Best Local Similarity 85.7%; Pred. No. 7.1;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ccgccttcaatctgctcatgcatcaccaccccc 35

|||||

Db 326 CTGCTTCAATCTGCTCATGCCCTACCCACCGGCC 360

RESULT 7

AI049127/c

LOCUS AI049127 288 bp mRNA EST 08-JUL-1998
DEFINITION ub35d04.r1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:1379719 5' similar to gb:X59869 TRANSCRIPTION FACTOR-7 (HUMAN
); gb:X61385 Mouse mRNA for T-cell specific transcription factor
(MOUSE);, mRNA sequence.

ACCESSION AI049127 GI:3297414
VERSION AI049127.1
KEYWORDS house mouse.
SOURCE Mus musculus

ORGANISM Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 288)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.

MG1:902187

Seq primer: -28ml3 rev2 Et from Amersham.

Location/Qualifiers

1..288

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:1379719"

/clone_lib="Soares_thymus_2NDMT"

/sex="male"

/tissue_type="Thymus"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Falima Bonaldo."

BASE COUNT 45 a 70 c 108 g 65 t

ORIGIN

Query Match 74.9%; Score 26.2; DB 10; Length 288;

Best Local Similarity 90.3%; Pred. No. 13;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgccttcaatctgctcatgcatcaccacc 31

|||||

Db 150 CTGCTTCAATCTGCTCATGCCCTACCCAC 120

RESULT 8

AA958693/c

LOCUS AA958693 330 bp mRNA EST 08-MAY-1998

DEFINITION v294h11.r1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:1344933 5' similar to gb:X59869 TRANSCRIPTION FACTOR-7 (HUMAN
); gb:X61385 Mouse mRNA for T-cell specific transcription factor
(MOUSE);, mRNA sequence.

ACCESSION AA958693

AA958693.1 GI:3124923

VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 330)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:693725
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 167.

FEATURES
source

1..330
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1344933"
/clone_lib="Soares_thymus_2NDMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
polylinker: Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

56 a 84 c 117 g 73 t
Query Match 74.9%; Score 26.2; DB 10; Length 330;
Best Local Similarity 90.3%; Pred. No. 13;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ccgccttcaatctgcctgctgattaccacc 61
Db 156 CTGCTTCAATCTGCTATGCCCTACCCACC 126

RESULT 9
LOCUS

BF538482 587 bp mRNA EST 11-DEC-2000
DEFINITION BF538482 Mus musculus cDNA clone IMAGE:4193012 5', mRNA sequence.
ACCESSION BF538482.1 GI:11625850
VERSION BF538482.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 587)

FEATURES
source

1..516
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM023b12_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9524 row: j column: 21
High quality sequence stop: 587.

FEATURES
source

1..587
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4193012"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="organ: salivary gland; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT
ORIGIN

146 a 199 c 136 g 106 t
Query Match 74.9%; Score 26.2; DB 11; Length 587;
Best Local Similarity 90.3%; Pred. No. 14;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ccgccttcaatctgcctgctgattaccacc 31
Db 7 CTGCTTCAATCTGCTATGCCCTACCCACC 37

RESULT 10
LOCUS

AV387748 516 bp mRNA EST 29-SEP-2000
DEFINITION AV387748 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii cDNA clone CM023b12_r, mRNA sequence.
ACCESSION AV387748
VERSION AV387748.1 GI:6541964
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 516)
Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S.
A large scale structural analysis of cDNAs in a unicellular green alga, Chlamydomonas reinhardtii. I. Generation of 3433 non-redundant expressed sequence tags
DNA Res. 6 (6), 369-373 (1999)
20152988
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakami@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.

JOURNAL
MEDLINE
COMMENT

FEATURES
source

1..516
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM023b12_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"


```

BASE COUNT      117 a      157 c      164 g      78 t
ORIGIN

Query Match      64.0%; Score 22.4; DB 10; Length 516;
Best Local Similarity 81.2%; Pred. No. 2.9e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ccgcttcaatctgctcatgcattaccacccc 32
|| ||||| ||||| ||| ||||| |||||
Db 283 CCTGCTCAAGTCTGATGAATTACCGACCC 314

RESULT 11
LOCUS      A1234477      465 bp      mRNA      EST      02-OCT-2000
DEFINITION EST380 Manduca sexta male antennae Uni-ZAP XR library Manduca sexta
          cDNA clone pMsmad135 5' similar to glyceraldehyde-3-phosphate
          dehydrogenase, mRNA sequence.
ACCESSION  A1234477
VERSION    A1234477.1 GI:3827995
KEYWORDS   EST.
SOURCE     tobacco hornworm.
ORGANISM   Manduca sexta
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
          ; Spingioidea; Spingidae; Spinginae; Manduca.
REFERENCE  1 (bases 1 to 465)
AUTHORS   Robertson,H.M., Martos,R., Sears,C., Todres,E.Z., Walden,K.O. and
          Nardi,J.B.
TITLE     Diversity of odorant binding proteins revealed by an expressed
          sequence tag project on male Manduca sexta moth antennae
JOURNAL    Insect Mol. Biol. 8, 501-518 (1999)
COMMENT    Contact: Robertson HM
          Department of Entomology
          University of Illinois at Urbana-Champaign
          505 S. Goodwin, Urbana, IL 61801, USA
          Tel: 217 333-0489
          Fax: 217 244 3499
          Email: hughrobe@uiuc.edu
          Insert Length: 1500 Std Error: 0.00
          Seq primer: T3
          High quality sequence stop: 300.
          Location/Qualifiers
FEATURES             source
     source           1..465
     /organism="Manduca sexta"
     /db_xref="taxon:7130"
     /clone="pMsmad135"
     /clone_lib="Manduca sexta male antennae Uni-ZAP XR
     library"
     /sex="male"
     /dev_stage="newly eclosed adults and pharate adults"
     /lab_host="Xl1 Blue MRF and SOLR"
     /note="Organ: antennae; Vector: Uni-ZAP XR; Site_1: EcoRI;
     Site_2: XhoI; The library was prepared by Stratagene using
     oligo-T priming and unidirectional cloning with an
     adaptor at the 5' end (GGCAGCAG) following the EcoRI
     site. The mRNA was prepared from antennae of late pupal
     and newly eclosed male moths. Clones were subcloned in
     vivo in mass into pBluescript maintained in SOLR cells
     for DNA sequencing."
BASE COUNT      97 a      147 c      110 g      111 t
ORIGIN

Query Match      63.4%; Score 22.2; DB 10; Length 465;
Best Local Similarity 77.1%; Pred. No. 3.4e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ccgcttcaatctgctcatgcattaccacccc 35
|| ||| ||||| || || ||| |||||
Db 320 CCTGCTCAATCTACTGGTGTATTCCCTACCCCCC 354

RESULT 12
LOCUS      BF279037      701 bp      mRNA      EST      07-MAR-2001
DEFINITION GA_Eb0036M14f Gossypium arboreum 7-10 dpa fiber Library Gossypium
          arboreum cDNA clone GA_EB0036M14f, mRNA sequence.
ACCESSION  BF279037
VERSION    BF279037.1 GI:11210107
KEYWORDS   EST.
SOURCE     Gossypium arboreum.
ORGANISM   Gossypium arboreum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE  1 (bases 1 to 701)
AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
          ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE     An integrated analysis of the genetics, development, and evolution
          of the cotton fiber
JOURNAL    Unpublished (2000)
COMMENT    Contact: Wing RA
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 7288
          Fax: 864 656 4293
          Email: rwing@clemson.edu
          Seq primer: TAATACGACTCACTATAGGG
          High quality sequence start: 147
          High quality sequence stop: 698.
          Location/Qualifiers
FEATURES             source
     source           1..701
     /organism="Gossypium arboreum"
     /strain="AKA"
     /cultivar="8400"
     /db_xref="taxon:29729"
     /clone="GA_Eb0036M14f"
     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
     /tissue_type="Fibers isolated from bolls harvested 7-10
     dpa"
     /lab_host="E. coli"
     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      189 a      175 c      112 g      207 t      18 others
ORIGIN

Query Match      62.3%; Score 21.8; DB 11; Length 701;
Best Local Similarity 78.8%; Pred. No. 4.9e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 cgccttcaatctgctcatgcattaccacccc 34
||||| || ||| ||||| || |||||
Db 244 CGCTTCCAACCTCTCTCGATCCTCATCACCCTC 276

RESULT 13
LOCUS      AU031552/c      558 bp      mRNA      EST      30-OCT-1998
DEFINITION AU031552 Rice cDNA from immature leaf including apical meristem
          Oryza sativa cDNA clone E61840_1A, mRNA sequence.
ACCESSION  AU031552
VERSION    AU031552.1 GI:3767442
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 558)
AUTHORS   Sasaki,T. and Yamamoto,K.
TITLE     Rice cDNA from immature leaf including apical meristem
          Unpublished (1997)
JOURNAL    Contact: Takuji Sasaki
COMMENT

```

8 days after inoculation with *Colletotrichum trifolii*"
 /lab_host="E. coli strain XL0LR"
 /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 cotyledons and primary leaves harvested 5 and 8 days after
 inoculation with *Colletotrichum trifolii*. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells. Note: EST may
 be of fungal origin."

BASE COUNT 176 a 118 c 110 g 208 t
 ORIGIN

Query Match 61.1%; Score 21.4; DB 11; Length 612;
 Best Local Similarity 80.6%; Pred. No. 6.7e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 4 cttcaaatctgctcattaccacccccc 34
 |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 21 CCTTTATCTACTCATGCAITTAACACACC 51

RESULT 15
 CNS003NE/c
 LOCUS
 DEFINITION
 Drosophila melanogaster genome survey sequence T7 end of BAC #
 BACR08115 of RPCT-98 library from *Drosophila melanogaster* (fruit
 fly), genomic survey sequence.

ACCESSION AL064726.1 GI:4942280
 VERSION AL064726
 KEYWORDS
 SOURCE fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 970)
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequences :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the *Drosophila*
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org/TheBDGP/Drosophila>
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammos in Pletier de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCT-98 and was constructed by partial
 EcoRI digestion of *Drosophila* DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the HDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..970
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCT-98"
 /clone="BACR08115"
 /note="end : T7"

BASE COUNT 243 a 189 c 243 g 223 t 72 others
 ORIGIN

Query Match 61.1%; Score 21.4; DB 13; Length 970;
 Best Local Similarity 80.6%; Pred. No. 6.9e+02;

	Matches	25;	Conservative	0;	Mismatches	6;	Indels	0;	Gaps	0;
Oy	1		ccgccttcaatctgctcattgattaccacc	31						
Db	353		CCGCCTTCGATATGCTCTCCAGTACCGACC	323						

Search completed: February 20, 2002, 13:52:34
Job time: 16223 sec